

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:23:59 ; Search time 8119.47 Seconds  
(without alignments)  
11509.077 Million cell updates/sec

Title: US-09-892-635A-44  
Perfect score: 2156  
Sequence: 1 ggaatcccaacttttaggaat.....agcctctcattcactatgg 2156

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: gb.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2125.6	98.6	4924	6	BD080859	Banana pr
c 2	185.2	8.6	298	8	AF399940	Musa acum
c 3	127.6	5.9	282699	8	AF506028	Poncirus
c 4	90.2	4.2	99606	8	AC116033	Medicago
c 5	87.8	4.1	86019	2	AC133571	Zea mays
c 6	87.8	4.1	121952	8	AC146745	Medicago
c 7	85.6	4.0	132605	8	AP004957	Lotus cor
c 8	83.8	3.9	109284	2	AC146910	Medicago
c 9	83.6	3.9	125883	8	AC124959	Medicago
c 10	83.2	3.9	141590	2	AC126788	Medicago
c 11	83	3.8	141590	2	AC126788	Medicago
c 12	81.2	3.8	100810	2	AF111804	Lycopersi
c 13	81	3.8	122497	8	AC087797	Oryza sat
c 14	81	3.8	157436	2	AC144740	Oryza sat
c 15	80.6	3.7	135301	8	AP003198	Oryza sat
c 16	80.6	3.7	142320	8	AP003199	Oryza sat
c 17	80.4	3.7	167484	2	AC145225	Zea mays
c 18	80.2	3.7	144000	8	AP005247	Oryza sat
c 19	79.8	3.7	88947	2	AP005463	Oryza sat
c 20	79.8	3.7	106632	8	AP004500	Lotus cor
c 21	78.8	3.7	129921	8	AC146568	Medicago
c 22	78.4	3.6	135983	2	AC144538	Medicago
c 23	78.4	3.6	136470	2	AC145222	Medicago
c 24	78	3.6	177140	2	AP005612	Oryza sat
c 25	77.8	3.6	87613	8	AP004857	Oryza sat
c 26	77.8	3.6	107580	8	AP005590	Oryza sat
c 27	77.6	3.6	133759	8	OSJUN00289	Oryza sat
c 28	77.4	3.6	118733	2	AP003885	Oryza sat
c 29	77.4	3.6	120419	8	AP005734	Oryza sat
c 30	77.4	3.6	143515	8	AP003252	Oryza sat
c 31	77.4	3.6	152973	8	AP004463	Oryza sat
c 32	77.2	3.6	146408	8	AC130610	Oryza sat
c 33	76.6	3.6	97579	2	AC146722	Medicago
c 34	76.4	3.5	74462	8	AC130731	Oryza sat
c 35	76.4	3.5	140553	8	AC091666	Oryza sat
c 36	76.4	3.5	161250	8	AC079634	Genomic S
c 37	76.4	3.5	300029	8	AE017080	Oryza sat
c 38	76.2	3.5	58293	8	AC096781	Oryza sat
c 39	76.2	3.5	162475	2	AC145272	Oryza sat
c 40	76.2	3.5	168267	8	AP003535	Oryza sat
c 41	76.2	3.5	199662	2	AC146338	Oryza sat
c 42	76.2	3.5	306161	8	AE017102	Oryza sat
c 43	76	3.5	110000	8	AC145127	Continuation (13 o
c 44	76	3.5	137015	8	AC113339	Genomic s
c 45	76	3.5	144778	8	OSJUN00146	AL662935 Oryza sat

ALIGNMENTS

RESULT 1  
BD080859  
LOCUS  
DEFINITION Banana proteins, DNA, and DNA regulatory elements associated with  
fruit development.  
ACCESSION BD080859  
VERSION BD080859.1 GI:22626462  
KEYWORDS JP 2001517446-A/9.  
SOURCE Musa acuminata  
ORGANISM Musa acuminata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;  
Musa.  
REFERENCE 1 (bases 1 to 4924)  
AUTHORS May, G. and Clendennen, S.

BD080859 4924 bp DNA linear PAT 27-AUG-2002  
Banana proteins, DNA, and DNA regulatory elements associated with  
fruit development.

TITLE Banana proteins, DNA, and DNA regulatory elements associated with fruit development  
JOURNAL Patent: JP 2001517446-A 9 OCT-2001;  
BOYCE THOMPSON INSTITUTE FOR PLANT RESEARCH INC  
COMMENT OS Musa acuminata (banana)  
PN JP 2001517446-A/9  
PD 09-OCT-2001  
PF 23-SEP-1998 JP 2000512960  
PR 25-SEP-1997 US 60/060062  
PI GREGORY MAY,STEPHANIE CLENDENEN  
PC C12N15/09,A01H5/00,C07K14/415,C07K14/42,C07K14/825,C12N5/10,C12N5/08,  
PC C12N9/10,C12N9/24,C12N15/00,C12N5/00  
CC Banana proteins, DNA, and DNA regulatory elements associated with fruit development  
CC development  
FH Key  
FT source

FEATURES  
source  
1. 4924  
Location/Qualifiers  
/organism="Musa acuminata"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4641"

ORIGIN  
Query Match 98.6%; Score 2125.6; DB 6; Length 4924;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2152; Conservative 0; Mismatches 0; Indels 5; Gaps 2;  
QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAA 60  
DB 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAA 60  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATGTAGTCAATCGGTGAAACCGGTGACATGTCTCGA 120  
DB 61 TCTTTACCAAGAGCTTTGAGTCCATGTAGTCAATCGGTGAAACCGGTGACATGTCTCGA 120  
QY 121 TGGACTCACTGGTTTCATTCGGAAGCTTCGAAAGCTTCGAAAGCTTCGAAAGCTTCGAA 180  
DB 121 TGGACTCACTGGTTTCATTCGGAAGCTTCGAAAGCTTCGAAAGCTTCGAAAGCTTCGAA 180  
QY 181 ATTCTTTCACTCGGTTGGTCTTCAATGAGTCACTCAAGAGTCTCCCAATATCAAAAG 240  
DB 181 ATTCTTTCACTCGGTTGGTCTTCAATGAGTCACTCAAGAGTCTCCCAATATCAAAAG 240  
QY 241 CCGAATCAAAATGAAATGATGTAATTCATTTTGTCTTAATGCAAAACAGGGCAT 300  
DB 241 CCGAATCAAAATGAAATGATGTAATTCATTTTGTCTTAATGCAAAACAGGGCAT 300  
QY 301 TCATAGCCTTTGTGTTTAAAGCAAAACATCTCTCCGATTCACTCCGATTCCGATTCCGTCATCG 360  
DB 301 TCATAGCCTTTGTGTTTAAAGCAAAACATCTCTCCGATTCACTCCGATTCCGTCATCG 360  
QY 361 GAAGAGAAAATTTTGAATCCATTTTCGACAAATAGACCAAGAGTCCGAAATCCATGGAAA 420  
DB 361 GAAGAGAAAATTTTGAATCCATTTTCGACAAATAGACCAAGAGTCCGAAATCCATGGAAA 420  
QY 421 TGAGAGAGATCCTCATATGAGTTTCCAAATCATGTAATTCGATCAATTAACATAGGTG 480  
DB 421 TGAGAGAGATCCTCATATGAGTTTCCAAATCATGTAATTCGATCAATTAACATAGGTG 480  
QY 481 GATGTGTAATGAATGACCTCATGCSCTATCTCTTGGGTATTAACCAAAATATGAGA 540  
DB 481 GATGTGTAATGAATGACCTCATGCSCTATCTCTTGGGTATTAACCAAAATATGAGA 540  
QY 541 GTGAGCCTTGCTCTGATACCAATTTTAGTATGATGAGTGGCACTAAGAGAGGGGGAGT 600  
DB 541 GTGAGCCTTGCTCTGATACCAATTTTAGTATGATGAGTGGCACTAAGAGAGGGGGAGT 600  
QY 601 GAATTAGTGCAGTGGATTAACACTATAAGTTTAAATGAATTCGTAAATACGAGAAGA 660  
DB 601 GAATTAGTGCAGTGGATTAACACTATAAGTTTAAATGAATTCGTAAATACGAGAAGA 660

QY 661 TTTGGTTTAAATAGTAACTTGGTAGATGAAAAACAAAAGTTAAACAGTAGTGTAAATAAC 720  
DB 661 TTTGGTTTAAATAGTAACTTGGTAGATGAAAAACAAAAGTTAAACAGTAGTGTAAATAAC 720  
QY 721 AATTTCCGGAAGTAAGAACTCACACATTCAAGGAACATACCAATTTAAAGTGGTTCGGT 780  
DB 721 AATTTCCGGAAGTAAGAACTCACACATTCAAGGAACATACCAATTTAAAGTGGTTCGGT 780  
QY 781 CAAAATGACCTTACATCCACTTTGTGAAGCTTTCTTGAAGAGGCTCCCAACTTCCACTAGC 840  
DB 781 CAAAATGACCTTACATCCACTTTGTGAAGCTTTCTTGAAGAGGCTCCCAACTTCCACTAGC 840  
QY 841 AAATCACTTTGAAGGGGAAAGCAAAATACCTCTCTTACNAGCTTTTACAATGGTTTCATC 900  
DB 841 AAATCACTTTGAAGGGGAAAGCAAAATACCTCTCTTACNAGCTTTTACAATGGTTTCATC 900  
QY 901 TCTTACAAATTTTCAACGAGAAAGAGGAGTGAACATGCAAGCAATGAAACCAAGACT 960  
DB 901 TCTTACAAATTTTCAACGAGAAAGAGGAGTGAACATGCAAGCAATGAAACCAAGACT 960  
QY 961 TGCTAAAGACTTTGCTAAGGCTTTTCTCAATCTATTGCTTCTCAAAAGTTGATTTCT 1020  
DB 961 TGCTAAAGACTTTGCTAAGGCTTTTCTCAATCTATTGCTTCTCAAAAGTTGATTTCT 1020  
QY 1021 CTGCTGAGAAATTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080  
DB 1021 CTGCTGAGAAATTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080  
QY 1081 TCGAATGCTTTGGGTTCCCGAGGTTGCGGTGCGCACCGCTGTCAAGTGTGACACTGG 1140  
DB 1081 TCGAATGCTTTGGGTTCCCGAGGTTGCGGTGCGCACCGCTGTCAAGTGTGACACTGG 1140  
QY 1141 ACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGTTGTTGGCGGTGCCACCGCTAG 1200  
DB 1141 ACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGTTGTTGGCGGTGCCACCGCTAG 1200  
QY 1201 ACTTTTTCAGTCACTGGTTGGATTCCAAACTTGACCCAAACCAAGTCCGAACTGGGTCC 1260  
DB 1201 ACTTTTTCAGTCACTGGTTGGATTCCAAACTTGACCCAAACCAAGTCCGAACTGGGTCC 1260  
QY 1261 AATTTGACCGGTAAACCGGATTATAGGATTAAACCTTAATCTCTAACTAATATGCAAA 1320  
DB 1261 AATTTGACCGGTAAACCGGATTATAGGATTAAACCTTAATCTCTAACTAATATGCAAA 1320  
QY 1321 CTACGCAACTGAAAATATAGTCTTAAGCAAGTTTTTAAACCGGCAAAAGTCCGAGTCTTCT 1380  
DB 1321 CTACGCAACTGAAAATATAGTCTTAAGCAAGTTTTTAAACCGGCAAAAGTCCGAGTCTTCT 1380  
QY 1381 CCGGCGATCTTTCCGCGAGCTTTCTGATATACCTTTGGATTCTTTCTAGCGGACTCCCTAGT 1440  
DB 1381 CCGGCGATCTTTCCGCGAGCTTTCTGATATACCTTTGGATTCTTTCTAGCGGACTCCCTAGT 1440  
QY 1441 AGGTCGCCGATCTTTGTGCGGAGTTTAGCGAGTACCGAACTCTCGGTGATCTCCGCAA 1500  
DB 1441 AGGTCGCCGATCTTTGTGCGGAGTTTAGCGAGTACCGAACTCTCGGTGATCTCCGCAA 1500  
QY 1501 ACCGCCGATGATCTCTTCCGCGAGCTTTCCGAAACTTTCGCAAGTCCCGGATTTCTTCTC 1560  
DB 1501 ACCGCCGATGATCTCTTCCGCGAGCTTTCCGAAACTTTCGCAAGTCCCGGATTTCTTCTC 1560  
QY 1561 GGTGTTCCGACAGCATCTCTACGAACTTCGGAATCTTGAATCTTGAATCTTGAATCTTGA 1620  
DB 1561 GGTGTTCCGACAGCATCTCTACGAACTTCGGAATCTTGAATCTTGAATCTTGAATCTTGA 1620  
QY 1621 ACTCCGCTAGGCTTTGCTTTATTTTTCAGGCTATCATAGTTAATCTTACATACATCTC 1680  
DB 1621 ACTCCGCTAGGCTTTGCTTTATTTTTCAGGCTATCATAGTTAATCTTACATACATCTC 1680  
QY 1681 AATTAATGATGATTAATTAACCCATCAATTTGATTTTATCATCATCAAAATTCGACATTC 1740  
DB 1681 AATTAATGATGATTAATTAACCCATCAATTTGATTTTATCATCATCAAAATTCGACATTC 1740

```

Qy 1741 AACAAACATCGTACTCAATAACCCATCAGCTATAGTTAGTGACTATCTACTGTGATC 1800
Db |||||
Qy 1741 AACAAACATCGTACTCAATAACCCATCAGCTATAGTTAGTGACTATCTACTGTGATC 1800
Db |||||
Qy 1801 CGTACGTGAAGTTAGCGAGTCATGATCAGGTGCTGTGCACCTTATTGGCCGAAACACGCTATC 1860
Db |||||
Qy 1801 CGTACGTGAAGTTAGCGAGTCATGATCAGGTGCTGTGCACCTTATTGGCCGAAACACGCTATC 1860
Db |||||
Qy 1861 CCTTATCAAAATCAGTCTTCTCAACTCTTCTAGACCTACCGCTCTCTTTTATTACTT 1920
Db |||||
Qy 1861 CCTTATCAAAATCAGTCTTCTCAACTCTTCTAGACCTACCGCTCTCTTTTATTACTT 1920
Db |||||
Qy 1921 TTGAAGAATTCAAATCAAAACAGATACAAATAACACGGTGAGACACTGTGACATGCTA 1980
Db |||||
Qy 1921 TTGAAGAATTCAAATCAAAACAGATACAAATAACACGGTGAGACACTGTGACATGCTA 1980
Db |||||
Qy 1981 GTCTCTGGAAGCAATTAATTCGGCATCCACAGACGTCGTCACTTCCATCACCACCTTT 2040
Db |||||
Qy 1981 GTCTCTGGAAGCAATTAATTCGGCATCCACAGACGTCGTCACTTCCATCACCACCTTT 2040
Db |||||
Qy 2041 TCCTACAT-ACCATGTCGCATGGCTTTGTTGATGACAGACACCAAGCTTGCCTTTGG 2099
Db |||||
Qy 2041 TCCTACAT-ACCATGTCGCATGGCTTTGTTGATGACAGACACCAAGCTTGCCTTTGG 2100
Db |||||
Qy 2100 TTGTGCTTAACAGAGAGAGAGAGACAGACACCGATAGCTCTCTCACTTCACTATGG 2156
Db |||||
Qy 2101 TTGTGCTTAAC-AGAGAGAGAGAGAGACAGACCGATAGCTCTCTCACTTCACTATGG 2153
Db |||||

RESULT 2
AF399940/c
LOCUS AF399940 298 bp DNA linear PLN 09-OCT-2003
DEFINITION Musa acuminata clone Radkal2 repeat sequence.
ACCESSION AF399940
VERSION AF399940.1 GI:15529477
KEYWORDS
SOURCE Musa acuminata
ORGANISM Musa acuminata
REFERENCE 1 (bases 1 to 298)
AUTHORS Valarik,M., Simkova,H., Hribova,E., Safar,J., Dolezelova,M. and Dolezel,J.
TITLE Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (Musa spp.)
JOURNAL Chromosome Res. 10 (2), 89-100 (2002)
MEDLINE 21989232
PUBMED 11993938
REFERENCE 2 (bases 1 to 298)
AUTHORS Valarik,M., Simkova,H., Hribova,E., Safar,J., Dolezelova,M. and Dolezel,J.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovska 6, Olomouc 772 00, Czech Republic
FEATURES
source Location/Qualifiers
1. .298
/organism="Musa acuminata"
/mol_type="genomic DNA"
/cultivar="Pisang Mas"
/db_xref="taxon:4641"
/clone="Radkal2"
repeat_region 1. .298
ORIGIN
Query Match 8.6%; Score 185.2; DB 8; Length 298;
Best Local Similarity 89.8%; Pred. No. 7e-36;
Matches 221; Conservative 0; Mismatches 23; Indels 2; Gaps 2;
Qy 6 CCAACTTTTAGGAATGATCTTAAATTTTAGTTTATAAGTTCAAAGTTAGAAAATCTTT 65
Db |||||
Qy 246 CCCAACTTTTGAAGGATCTTAAGATTTTAGTGACTAGTTCAAAATTAGAAACT-TTT 188
Db |||||

```

```

Qy 66 ACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAC-GGTGTACATGTCTCGATGGA 124
Db |||||
Qy 187 ACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAACCGGTGTACATGTCTCGATGGA 128
Db |||||
Qy 125 CTCACTTGGTTTCATTTCGGAAGAGTTTCAAGAGTGCATGAAGATATTGATTTTGGATTC 184
Db |||||
Qy 127 CTCACTTGGTTTATCCGGAAGAGTTTCTGAAGATGCAAGATATTGATTTTGGACTC 68
Db |||||
Qy 185 TTTCACTCGGTTGGTGCCTTTCATGAGTACCTCAAGAGTCTCTCAATATCAAAAGCCGA 244
Db |||||
Qy 67 TTTCACTCGGCTAGTGCCTTTCATGAGTACCTCTAGAGTTCTCCAAATATCAAAAGCCGA 8
Db |||||
Qy 245 ATCACA 250
Db |||||
7 GTCACA 2

RESULT 3
AF506028/c
LOCUS AF506028 282699 bp DNA linear PLN 21-FEB-2003
DEFINITION Poncirus trifoliata citrus tristeza virus resistance gene locus, complete sequence.
ACCESSION AF506028
VERSION AF506028.1 GI:24461847
KEYWORDS
SOURCE Poncirus trifoliata
ORGANISM Poncirus trifoliata
REFERENCE 1 (bases 1 to 282699)
AUTHORS Yang,Z.-N., Ye,X.-R., Choi,S., Molina,J., Moonan,F., Wing,R.A., Roose,M.B. and Mirkov,T.E.
TITLE Construction of a 1.2-Mb contig including the citrus tristeza virus resistance gene locus using a bacterial artificial chromosome library of Poncirus trifoliata (L.) Raf
JOURNAL Genome 44 (3), 382-393 (2001)
MEDLINE 21337599
PUBMED 11444697
REFERENCE 2 (bases 1 to 282699)
AUTHORS Yang,Z.-N., Ye,X.-R., Molina,J., Roose,M.L. and Mirkov,T.E.
TITLE Sequence Analysis of a 282-kilobase Region Surrounding the Citrus Tristeza Virus Resistance Gene (Ctv) Locus in Poncirus trifoliata L. Raf
JOURNAL Plant Physiol. 131 (2), 482-492 (2003)
MEDLINE 22474771
PUBMED 12586873
REFERENCE 3 (bases 1 to 282699)
AUTHORS Yang,Z.-N., Ye,X.-R., Molina,J., Roose,M.L. and Mirkov,T.E.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2002) Dept. of Plant Pathology and Microbiology, The Texas A&M University Agricultural Experiment Station, 2415 E. Hwy 83, Weslaco, TX 78596, USA
FEATURES
source Location/Qualifiers
1. .282699
/organism="Poncirus trifoliata"
/mol_type="genomic DNA"
/cultivar="Pomeroy"
/db_xref="taxon:37690"
/notes="derived from BAC clones 27A14, 20J24, 83D17, and 84F5"
<306. .>5092
/genes="CTV.1"
join(<306. .377,493. .534,931. .972,2377. .2499,2870. .2926, 3477. .3614,4240. .4281,4312. .4379,5057. .>5092)
/genes="CTV.1"
/product="CTV.1"
join(<306. .377,493. .534,931. .972,2377. .2499,2870. .2926, 3477. .3614,4240. .4281,4312. .4379,5057. .5092)
/genes="CTV.1"
/notes="similar to F-box protein"
/codon_start=3

```

```

/product="CTV.1"
/protein_id="AAN62335.1"
/db_xref="GI:24461848"
/translation="IGSLFYLNLRQCQLSDGCKFNLGNEITDECLVVHLSDCGI
DEGLVNYLFTGIDSGLRKAGLSLKLNLDAQITDTGLAALSAHIDLFGARI
TDSGAAYLRNKLRLIEICGGGUTDAGVKHKOLSLALINLQNCNLTKTLEYVS
NSRITSAGLRPLESCKVTANDIKRLQSRDLPNLFLTPHTTSG"
repeat_region
4157..4166
/rpt_unit="at"
repeat_region
6640..6651
/rpt_unit="ta"
gene
complement(<7554..>15436)
/gene="CTV.2"
CDS
complement(join(<7554..7764,8721..8887,8973..9161,
9491..9586,9671..9799,9881..10060,10225..10335,
10762..10908,11100..11195,11277..11435,11518..11752,
11882..12009,12143..12238,12326..12421,12525..12598,
12686..12755,12863..12972,13107..13272,13405..13514,
13627..13753,13839..14016,14547..14681,14780..14893,
14993..15172,15357..>15436))
/gene="CTV.2"
/product="CTV.2"
complement(join(7554..7764,8721..8887,8973..9161,
9491..9586,9671..9799,9881..10060,10225..10335,
10762..10908,11100..11195,11277..11435,11518..11752,
11882..12009,12143..12238,12326..12421,12525..12598,
12686..12755,12863..12972,13107..13272,13405..13514,
13627..13753,13839..14016,14547..14681,14780..14893,
14993..15172,15357..15436))
/gene="CTV.2"
/translation="WSLSRELVLIIQLFLDEERFKETVHKLQESGFFFMKYPFDE
VHNGWDDVELKSGFTKVDNRYSMKIFPEIRKQYLEALDKHDAKAVELIVKDLK
VFSTFNEBELKEITQLLTLENFRENELQSKYDTKSARSIMLVKLIENPLPRDK
LOFNLRKSNLWQHQCCKNRPNDIKLFDVHTCGONGARAPSPANNELLGSLPK
AGVPPFGAGPFPPTAPVTPPLAGMNSNPVTTHPAVGGAGLGSPIPAALKH
PRTPPTNFSVDPSGSDHLSKRTPIGISDEINLPVNLVPSFTGHSQAFAPED
LPKVTITLQNGSSPMSMDHPVQTLILLVNTGVNVDIGLWEGVSRERLVRNFKVWDL
GCSMPTQALVAPKPGVSNRVIWSPDGLFGVAVSRHVOIYSYHGDEVROHLEID
AHVGVNDIAFSPHNKOLCVITCGDDKTIKVMDATNGAKQVIFSGHEAPVYSCPHKH
ENIQIFSTALDGKIKAWLYNLSRVDYERAPGRWCTNMSADGTRILFSCGTSKDG
SFIVWNESEGAVKRTYQGFRRKSLVVVQFDTTKNRFIAAGDDFSIKPMDMSYQLLT
SIDAGGLPASPRIRFNKDGCLAVTNDNGIKILATSDGIRLLRTFENLAYDSRTS
ENSKPTTSPISAAAAAATSGALDRAASWYSIFPMNGDVRSADVPRIITEENDKS
KWKLTELSPNOCRSRLPENLRATKISRLIFTNSGNAIILASNAIHLLWKWQRT
RNSGKATASVAPOLKOPPGSIMTNDVDSNPREAVPCFALSKNDSVWMSASGKIS
LFNMTEKTATFNPPPAATFLAFHPQDNNIIAIGMDDSSIQIYNNRVDEVKSLKG
HSKRITGLAFSHALNVLVSSGADSQLCWGSDGWEKQKNRFLQIPTRGTPAQSDTRY
QHGDQTHFLVHETQAIETFTKLECVKQWPRESSAPITATFSCDSQLVYACFLD
ATCVFCAAKLKRINPSAYLPAGVSSNVHPLVIAHPQEPNEFALGLSDGGVHVF
EPLESECKGWGPPVPVDNGSTSSMEATPVGSGSDQQR"
repeat_region
15984..15993
/rpt_unit="tc"
repeat_region
19033..19050
/rpt_unit="aat"
repeat_region
23206..23220
/rpt_unit="aat"
repeat_region
23248..23262
/rpt_unit="aat"
repeat_region
26450..26459
/rpt_unit="at"
repeat_region
36059..36074
/rpt_unit="at"
repeat_region
37008..37031
/rpt_unit="aat"
repeat_region
37044..37053
/rpt_unit="ag"
repeat_region
37140..37155
```

```

/rpt_unit="tc"
38595..38604
/rpt_unit="at"
<39061..>40533
/gene="CTV.3"
join(<39061..40137,40222..>40533)
/gene="CTV.3"
/product="CTV.3"
join(39061..40137,40222..40533)
/gene="CTV.3"
/notes="similar to transmembrane amino acid transporter
protein"
/product="CTV.3"
/protein_id="AAN62337.1"
/db_xref="GI:24461850"
/translation="MDSNYTVIPKNSFLEQLVYONDQNPSPRINSHVKIOTFDBEHDA
DSKNFVDHNNQDDHDDXPLISAKNMGSGVAGVFNLTITVIGAGIMALPAAMKV
LGLAVGVFLIIFMGILSEVVELLVRFLCKATSYGEVQYALGRPAKLSEICIL
NNAAGLVVLLIIVGDVMSGSLHHTGVFQWLKGMWDHRKLLIILVVFPLAPLCAJD
RIELSMSAASVALAVFVVVCFPIAFIKLVEGLDPPRMSPFQSKKAILDLVVI
PIMTNAVCHFNQVPIYNELEGRTPQNMHVGRUTTLICIVVSSSTAISGYLLFGKDT
ESDVLTFDKDLGIGYSTALNIVRGSAPLLESKRSLATGLVLLTYFGSTIMPSI
WTAPKFGATTAVSLGFIFFPLVALRLKEGPGSLWEKFLSGLMLVLAIVVSFVGVN
GNIYSLKSD"
<42823..>45495
/gene="CTV.4"
/notes="disease resistance gene R1"
<42823..>45495
/gene="CTV.4"
/product="NBS-LRR type disease resistance protein"
42823..45495
/gene="CTV.4"
/codon_start=1
/product="NBS-LRR type disease resistance protein"
/db_xref="GI:24461861"
/translation="MGNICOLISCDGAFNNRCLDFLGKAAVINSIQDNLVALDTL
EKLIAAKDMRRVNDABEQMRRLDQVQVWVSEVTEADAPIGDGTQIEIKLCL
GYCSKCKSKYFGKQVAKRLDKITMGEGVEFVADKVPFPADEPTEPTVVLG
OSQLEEVCLVEPVGIVGLYGMGSKTTLTHINNKFLGSPFNLDLVLVVSKD
LRLESIQEVIGKEIGLLNDWKSRRIEQKALDIPRI LRGNFVLLDDIQWRVLAJK
GILPNQTSASKVVFTRSEEVCLGMAMKFKVECLSGNDWELFRQKVGEEITLNC
HHDILEAQTVTKECGGLPLALITIGRAMACKTPEEWSVAIOVLRYSOFFPGLNE
VYPLKSYNLDNDTIRSCLLYCCLYPECCISKENIVDCWIGEGLLNGSVTIGSHE
QGVHVGILVHSCLEEVDEVMHDVIRDMALMACDAKKEILVYAGAGUREA
FDVIEWKLHSLMENQIENLSEVPTCPHLLTLFNSDDILMEINSDFLQSMRLKV
LNLRYMGLLVLPLIGISKLVSEYLDLSLISIPBELKALVNLKCLNLYTGLLKL
IPLQILNFGRLHLVRMFGNAYFSYGNVPTESVLFGGELLVEBLLGKHLEVLSLTL
GSRALQSFLTSHMLRSCTRAMLLODFQGSTSDVDSGLADLKRLLKRLRISPCYELVEL
KIDYAGVQVRYGHSLOSFBVNVCSKLKDLTLVLINLKSIEVTDCEAMBEIISVE
FAGNPAFLQYIGLIGNLNLKSIYWKPLFPFPCLEETVSDCYELKLLPUDSNAKE
HKIVIRGAANWRNLQWEDEATQNAFLSCFQSLA"
/gene="CTV.5"
join(<46474..46761,48604..49006,49060..49234,49534..49620,
50490..50665,51207..51670)
/product="monosaccharide transport protein"
join(46474..46761,48604..49006,49060..49234,49534..49620,
50490..50665,51207..51670)
/codon_start=1
/product="monosaccharide transport protein"
/protein_id="AAN62339.1"
/db_xref="GI:24461851"
/translation="MIESGAYNNYTLVQDPELLSRKRSALATGLLLVLIVPGSTMIPN
IWTAPKFGATTAVSLGFIFFPLVALRLKEGPGSLGKFLSGLMLVLAIGLLTH
SLIAKAVAAACIGAKFGIDGNPGELPKWYAIYVVLFPYIYVAGVAGPGPLGWLVPSE
ILPLKIRSAQSVNVSNMCLTFLVAQVFLNKLCHLKEGLLFPFGFVWMSFTYFPF
LAETKGIPIEMDWLTNDCTDNDENTTLHLKLOLHLKLLRLSRSPFLQDLRSQKQITG
AGGFGKNPMLNGLRGETIALLRDRAFFLNDKLLVGMRLCNVGPSSRKNLTLYSQP
EAIFPFGCELNRRGQSGSGENSSDGTETPGEKGGNEETIQSGLDQDHLHTRFV
TPRRLFFTLHVSGRHLFVYLTISMSRSDSWGLDDEPLVKVLEAFNAFNCDDL
SPTSLVBSRLVDVAGMDNKIPSAMSRKRAQVSGESSIAPFSQKKSNTGFSKTSAPAL
```



## LN1VNKEASGVRLHGNCCITE"

Query Match 5.9%; Score 127.6; DB 8; Length 282699;  
 Best Local Similarity 56.4%; Pred. No. 3.6e-21;  
 Matches 298; Conservative 0; Mismatches 199; Indels 31; Gaps 2;

QY 6 CCAACTTTTGGAGTGGATCTTAAATTTTGGTTTATTAAGTTTCAAGTTTGAAGAAATCTTT 65  
 DB 72738 CCACTCTTTAGTAGAGATCTAATGATTTTCTTGACITTTTCGCTAATTTGAAGAAGCTTT 72679

QY 66 ACCAAGAGCTTTGAGTCCATGATGACATCCGCGTGAAC-GGTGACATGCTCCGATGGA 124  
 DB 72678 TCCTAGGCTCCTAGAGCTGTTCACTATGTCGGTGAATCTAGTATACATAGATACACATT 72619

QY 125 CTCACTTGGTTTCACTCGGAAAGTTTCAAGAGAGTGCATAGATATATGTTTGGATTC 184  
 DB 72618 TTCACTTTTGTTCCTATTTGAAACAATTCGTATGTCGAGTGTATCTACTTATTTTAGATTC 72559

QY 185 TTTCACTCGGTTGGTGGCTTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCGGA 244  
 DB 72558 TTTCACTTGGTTGTACTCTATAGACAACCTCAAGTTTGTGCCAAATCTCAITAGCACT 72499

QY 245 ATCACAAAATTTGAATGATGATGATTCATTTTGTCTTAATGCAAAACAGGGCAATTCAT 304  
 DB 72498 TTGCAACTAGACACTCTATGAAACTCTTACTTATCTAGTGCACAGAACAGGCAATTCAT 72439

QY 305 AGCTTTTGTCTTTAAAGCAAAACATTTCTCTCGATTCATCCATTCGCTATCGGAAG 364  
 DB 72438 GGCCTTGGAAATTTAGAGAAGCTTTCTCTTATCGAATTCATCCCAATCCCGTGAAGGTTT 72379

QY 365 AGAAAAATTTTGAAT-----CCATTTTCGACAAT 394  
 DB 72378 TGGATATCTTCTCCTACTTCATTTTGTAGTCAAGGATCAAGGGCCATCAITAAACAT 72319

QY 395 AGACCAAGTCGAAATCCATGAGAAATGAGAAAGATCCTCATATGATTTTCCAAATACAT 454  
 DB 72318 TTCCCAATCTCGTAATCTAAGGCTTGCAGATTAATCTCATCTAGTTTCCAAATATGG 72259

QY 455 GTAATTCGATCTTAACATAGTGTGATGTAATGAATGACCTC 502  
 DB 72258 GTAATCATTACCAATCAAGTAAGGGGTCTAGTGTGGAGTGTCTCTTC 72211

RESULT 4  
 AC116033/c  
 LOCUS AC116033 99606 bp DNA linear PLN 04-OCT-2003  
 DEFINITION Zea mays genomic clone ZM15C05 sequence, complete sequence.  
 ACCESSION AC116033  
 VERSION AC116033.3 GI:37514986  
 KEYWORDS HTG.

SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 99606)  
 Nagaki,K., Song,J., Stupar,R.M., Parokorny,A.S., Yuan,Q.,  
 Ouyang,S., Liu,J., Hsiao,J., Jones,K.M., Dawe,K., Buell,R. and  
 Jiang,J.

TITLE Molecular and cytological analyses of large tracks of centromeric  
 DNA reveal the structure and evolutionary dynamics of maize  
 centromeres

JOURNAL Genetics 163 (2), 759-770 (2003)

MEDLINE 22505408

PUBMED 12618412

REFERENCE 2 (bases 1 to 99606)

AUTHORS Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,  
 Gansberger,K. and Buell,R.

TITLE Direct Submission

JOURNAL Submitted (23-MAR-2002) University of Wisconsin, Department of

REFERENCE 3 (bases 1 to 99606)

AUTHORS Horticulture, Madison, WI 53706, USA

REFERENCE 3 (bases 1 to 99606)

AUTHORS Jiang,J.

TITLE Direct Submission  
 JOURNAL Submitted (21-NOV-2002) University of Wisconsin, Department of  
 Horticulture, Madison, WI 53706, USA  
 REFERENCE 4 (bases 1 to 99606)  
 AUTHORS Jiang,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-NOV-2002) University of Wisconsin, Department of  
 Horticulture, Madison, WI 53706, USA  
 REFERENCE 5 (bases 1 to 99606)  
 AUTHORS Jiang,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-DEC-2002) University of Wisconsin, Department of  
 Horticulture, Madison, WI 53706, USA  
 REFERENCE 6 (bases 1 to 99606)  
 AUTHORS Jiang,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-OCT-2003) University of Wisconsin, Department of  
 Horticulture, Madison, WI 53706, USA  
 COMMENT On Oct 4, 2003 this sequence version replaced gi:25141017.  
 FEATURES  
 source  
 1. 99606  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4577"  
 /chromosome="unknown"  
 /clone="ZM15C05"  
 /note="The BAC contains DNA sequences associated with  
 maize centromeres."

ORIGIN  
 Query Match 4.2%; Score 90.2; DB 8; Length 99606;  
 Best Local Similarity 51.3%; Pred. No. 1.1e-11;  
 Matches 254; Conservative 0; Mismatches 228; Indels 7; Gaps 2;

QY 4 TCCCAACTTTTAGGAATGGAATCTTAAATTTTGGTTTATTAAGTTTCAAGTTTGAAGAAATCT 63  
 DB 33123 TCATATTTCTTCGGAAGTGAATCTAAGAACTTGTGTGAAATCCCATCCGGTACATTA 33064

QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCCGCGTGAACGGT-GTACATGCTCCGATG 122  
 DB 33063 AAACCAAGTCTTTGAGTTCATTTACTTACTTCTCATTCACCGGATGAACATATCGGATACA 33004

QY 123 GACTCACTGGTTTTCATTCGGAAGATTCGGAAGAGTGCATAGAATATTGATTTTGAT 182  
 DB 33003 CTCTCATCTTTTCATATAATAAATTTGCTCAAACTTCCCTTTGCACATATAGTTTGGCA 32944

QY 183 TCTTTCACCTGGTTGGTGGCTTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAGCC 242  
 DB 32943 CTCTTCACAAATTCGATTTCTTCTTCATGAAATTTCCATTTATTTTATCCAAATCTCATGAGCG 32884

QY 243 GAATCAAAATTCGAAATGATTTGAATTTCAATTTTGTCTAATGCACAAAACAGGGCATTC 302  
 DB 32883 GTTGTGAGATTTCTTGATACGATTTGAACCCATTTATATCAAGAGCATCATAAAGACATTC 32824

QY 303 ATAGCCCTTTGTGTTTAAAGCAAAACATTTCTCTCCGATTTCACTCCCATTCGCTCATCGGA 362  
 DB 32823 TTGGCTTGATCAATTTGTGAGGATATTTCTCATTTATCACTAACAGATGGTTTCATTTCTTC 32764

QY 363 AGAGAAAATTTTGAATCCATTTTCGCAATAGACAAAGCTCGAAATCCATCGGAATG 422  
 DB 32763 AACACAA-----CAATCCATCCCTGACAAATTTGGCCAAATTTTCCACCCATTTGCTCTA 32710

QY 423 AGGAAGATCTCATATGATTTTCCAAATACATGTAATTCGACTCATTAACATAGTGGGA 482  
 DB 32709 AGATGATTCGACATTTCTTATTTTCCCAATATCATATAATTTGTTCCATCAAAAGTCGGTGGC 32650

QY 483 TGTGTATG 491  
 DB 32649 TTTCCCAATG 32641

RESULT 5  
 AC133571/c

```

LOCUS      AC133571      86019 bp      DNA      linear      HTG 05-NOV-2003
DEFINITION Medicago truncatula clone mth2-10h12, WORKING DRAFT SEQUENCE, 2
ordered pieces.
ACCESSION  AC133571
VERSION    HTG; HTGS PHASE2; HTGS DRAFT.
KEYWORDS   Medicago truncatula
SOURCE     Medicago truncatula
ORGANISM   Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE  1 (bases 1 to 86019)
AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE     Medicago truncatula BAC Clone mth2-10h12
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 86019)
AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (14-SEP-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE  3 (bases 1 to 86019)
AUTHORS   Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (05-NOV-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT   On Nov 5, 2003 this sequence version replaced gi:380441179.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 23532: contig of 23532 bp in length
* 23533 23632: gap of unknown length
* 23633 86019: contig of 62387 bp in length.
* Location/Qualifiers
* 1..86019
* /organism="Medicago truncatula"
* /mol_type="genomic DNA"
* /db_xref="taxon:3880"
* /clone="mth2-10h12"
* /clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match 4.1%; Score 87.8; DB 2; Length 86019;
Best Local Similarity 51.5%; Pred. No. 4.5e-11;
Matches 226; Conservative 0; Mismatches 212; Indels 1; Gaps 1;

QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTGTATATAAGTTCAAAGTTAGAAAA 60
Db 56956 GGTCTCCATTGAGTTGGAGACATCTTAGTATTTTCTATCTCTCGTGGTGAATAA 56897
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTGATGATCCGTTGAAACG-GTGTACATGTCCTCG 119
Db 56896 GTTTTCCAGTGATCTTAGTTCACTTACTATAGATGTGAATCTAGATACATTTTCATCA 56837
QY 120 ATGGACTCACTGGTTTCATTCGGAAAGTTCCGAAGAGTGCATAGATATGATTTTG 179
Db 56836 ATGTTTTCATCTTCAATCATTTTCAAAGATTTCAAACCTTTCTCACCAATGTCATTTCT 56777

```

---

```

QY 180 GATTTCTTCACTCGTTGGTGCCTTCATGTAGTGACCTCAAGAGTCCCTCCRAATATCAAAA 239
Db 56776 GTTTCTTTAACAATGGCTTGTTTCCCTTCATGATGGAGACTTTGAGAGATATCCACACCTTTT 56717
QY 240 GCCGAATCACAAATTTGAAATGTGATTGAATTCATTTTGTGCTAATGCACAAAACAGGGCA 299
Db 56716 GCATATCACATTCATCACTCTTTCTACTTTTCCATGCTTAGTGACATGATAGAAG 56657
QY 300 TTCATAGCTTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATCGCTCATC 359
Db 56556 AGTCGAGCTTTAGAATTTAAAGATATTTTCTTTATCTTCTTTTGTCCATGATGCTTCT 56597
QY 360 GGAAGAGAAATTTTGAATCCATTTTCGACATAGACCAAGACTCGAAATCCATGAA 419
Db 56596 GGTGTAATTACTTTTGTGTTTGTAGTATATATCATCTGTGTCATGTGTACATGATCCCATCT 56537
QY 420 ATGAGGAAGATCCCTCATAT 438
Db 56536 GTTATTATTATTCACATAT 56518

RESULT 6
LOCUS    AC146745      121952 bp      DNA      linear      PLN 09-DEC-2003
DEFINITION Medicago truncatula clone mth2-53p19, complete sequence.
ACCESSION AC146745
VERSION    AC146745.17 GI:39573788
KEYWORDS   HTG.
SOURCE     Medicago truncatula (barrel medic)
ORGANISM   Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE  1 (bases 1 to 121952)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLE     Medicago truncatula BAC Clone mth2-53p19
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 121952)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (24-SEP-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE  3 (bases 1 to 121952)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (06-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE  4 (bases 1 to 121952)
AUTHORS   Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
TITLE     Direct Submission
JOURNAL   Submitted (09-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT   On Dec 9, 2003 this sequence version replaced gi:39229182.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
Location/Qualifiers
1..121952
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-53p19"

FEATURES
source

```

ORIGIN /clone\_lib="Medicago truncatula BAC library H2"

Query Match 4.1%; Score 87.8; DB 8; Length 121952;  
Best Local Similarity 55.0%; Pred. No. 4.5e-11;  
Matches 193; Conservative 0; Mismatches 157; Indels 1; Gaps 1;

QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTACGTTTAAAGTTCAAGTTAGAAAA 60  
|||  
Db 94945 GGCTCTCATGCTTGAAGACATCTTAAATTTTCTTATTCGATCATGAGTGGTATA 95004  
|||  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATGATCAGATCCGTGAAC- GGTGTACATGTCCTCCG 119  
|||  
Db 95005 GTTTTCCAGTGATCTTAATTCATTACATAGATGATGAACTGGAGTACATTTATCA 95064  
|||  
QY 120 ATGAGTCACTTGGTTTCAATCGGAAAGTTCGAAGAGTGCATAGAATATTGATTTTG 179  
|||  
Db 95065 ATGTTCTCATCTTGGTTTCAATTCGAAGATTTCAAAATTTTCTTACGCCAATATCTATT 95124  
|||  
QY 180 GATTTCTTCACTCGGTGGTGGCTTCATGATGACCTCAAGAGTCCCTCAAAATATCAAA 239  
|||  
Db 95125 GTTTCCTTACGGGTAGTTCCTCATGATGATTTTATAGAGTGCCTCAAACTCTTT 95184  
|||  
QY 240 GCCGAATCAAAATGAAATGATGAAATTCATTTTGTCTAATGCACAAACAGGGCA 299  
|||  
Db 95185 GCAGTAACATTCATCACTCTTCACTTCTTCCATGCTTGTGGCATGATAGAAG 95244  
|||  
QY 300 TTCAATAGCCTTTGTTTAAAGCAAAACATCTTCTCCGATCAATCCCAT 350  
|||  
Db 95245 AGTCGAGCTTTGGAGTTAGAGTACTTTTCTTATCTTCTTTTGACCAT 95295  
|||

RESULT 7  
AP004957  
LOCUS 132605 bp DNA linear PLN 22-JUL-2003  
DEFINITION Lotus corniculatus var. japonicus genomic DNA, chromosome 2,  
clone: LjT34H20, TM0134, complete sequence.

ACCESSION AP004957  
VERSION AP004957.1 GI:21907975  
KEYWORDS HTG.  
SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae;  
Lotus.

REFERENCE 1  
Kaneko, T., Nakamura, Y., Asamizu, E., Kato, T., Sato, S. and Tabata, S.  
Structural Analysis of a Lotus japonicus Genome. I. Sequence  
Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb  
Regions of the Genome  
Unpublished  
2 (bases 1 to 132605)  
Nakamura, Y.  
Direct Submission  
Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research; 1532-3, Yana,  
Kisarazu, Chiba 252-0812, Japan (E-mail: ynakazusa.or.jp.  
URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,  
Fax: 81-438-52-3934)  
LOCATION/Qualifiers  
1. 132605  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="genomic DNA"  
/variety="japonicus"  
/db\_xref="taxon:34305"  
/chromosome="2"  
/clone="LjT34H20"  
/clone\_lib="LjT library"  
/note="TAC clone: TM0134-synonym: Lotus japonicus"

FEATURES  
source

ORIGIN

Query Match 4.0%; Score 85.6; DB 8; Length 132605;

Best Local Similarity 50.8%; Pred. No. 1.6e-10;  
Matches 268; Conservative 1; Mismatches 240; Indels 19; Gaps 2;

QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTACGTTTAAAGTTCAAGTTAGAAAA 60  
|||  
Db 112738 GGATCCAGCTTTTTCAGGAAGACCTCTAAGAATCCTCATAACATGATCACCAGTTGTGTA 112797  
|||  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTTGATGA-CATCCGTGAAACGGTGTACATGTCCTCCG 119  
|||  
Db 112798 CTTTATTGAGGGGTCTTAATTCAGCAATAGTAATCGAAATCTGGAGAACATGTCTTCA 112857  
|||  
QY 120 ATGAGTCACTTGGTTTCAATTCGAAAAAGTTCGAAAGAGTGCATAAGATATTGATTTTG 179  
|||  
Db 112858 ATGAGTCACTGATTCATTTGGAAGGATTCATATCTCTGATCAGAGCAACGCCCTTT 112917  
|||  
QY 180 GATCTTTCACTCGGTGGTGGCTTCATGAGTGCCTCAAGAGTCCCTCAAAATATCAAA 239  
|||  
Db 112918 GACTCTTTCACTCTTTGTTTCCATCATGGGTCTCTTCAAGGAATCAAGATACCCCTTA 112977  
|||  
QY 240 GCCGAATCAAAATGAAATGATGAAATTCATTTTGTCTAATGCACAAACAGGGCA 299  
|||  
Db 112978 GCATCTCAGATCTGTAATCTTCTGGTATCTTCAATAGAAATAGCATGAGAAGATA 113037  
|||  
QY 300 TTCAATAGCCTTTGTTTAAAGCAAAACATCTCTTCCGATTCATCCCATCGCTCATC 359  
|||  
Db 113038 GTTCTAGCTTTATGATGCGAGTGAAGTAAAGCTCTTCTGCTCGAGTCACTCTGTTTG 113097  
|||  
QY 360 G-----GAAGAGAAAAATTTTGAATCCATTTTCGACAAATAGACCAA 401  
|||  
Db 113098 GAGATCTTCTGTCATCACCATCACTGGAAGCTCGTAACCATCCACTATAATATCCAG 113157  
|||  
QY 402 AGCTCGAAATCCATGGAATGAGGAAGATCCTCATATGATGATTTTCCATACATGTAATTC 461  
|||  
Db 113158 AGATCTGCATCAAGGCCCGAGAAAAAGCTTTCAATCTGCTTCCAGTACTCAAAATCTT 113217  
|||  
QY 462 GACTCAATTAACATAGTGTGATGTAAATGAATGACCCCTCATGCSCT 509  
|||  
Db 113218 TGACCATCAACATAGAGGCTTTGCACGTAGACATCTTTTGGCGT 113265  
|||

RESULT 8  
AC146910/c

LOCUS 109284 bp DNA linear HTG 06-DEC-2003  
DEFINITION Medicago truncatula clone mth2-59121, WORKING DRAFT SEQUENCE, 2  
unordered pieces.

ACCESSION AC146910  
VERSION AC146910.5 GI:39228867  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

REFERENCE 1 (bases 1 to 109284)  
AUTHORS Lin S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
and Roe, B. A.

TITLE Medicago truncatula BAC Clone mth2-59121  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 109284)  
AUTHORS Lin S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
and Roe, B. A.

TITLE Direct Submission  
JOURNAL Submitted (21-OCT-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

REFERENCE 3 (bases 1 to 109284)  
AUTHORS Lin S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
and Roe, B. A.

TITLE Direct Submission  
JOURNAL Submitted (06-DEC-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA

COMMENT On Dec 6, 2003 this sequence version replaced gi:38371848.  
 ----- Genome Center  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 30237: contig of 30237 bp in length  
 \* 30238 30337: gap of unknown length  
 \* 30338 109284: contig of 78947 bp in length.  
 FEATURES  
 source  
 1. 109284  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-59121"  
 /clone\_lib="Medicago truncatula BAC library H2"  
 ORIGIN  
 Query Match 3.9%; Score 83.8; DB 2; Length 109284;  
 Best Local Similarity 54.5%; Pred. NO. 4.7e-10;  
 Matches 189; Conservative 0; Mismatches 157; Indels 1; Gaps 1;  
 QY 7 CAACCTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAAGTTAGAAATCTTTA 66  
 Db 1100 CAAGATCTGGAGACATCTTAAATTTTCTTACTCTTCATGATGTAATCTCTT 1041  
 QY 67 CCAAGAGCTTTGAGTCCATGATGACATCCGTGAAACGGTG-TACATGTCCTCCGATGCAC 125  
 Db 1040 TCAAGAGAGCGTAATCATTAATGATAGTAAACCTTCCATACATTTTCATCTATGTT 981  
 QY 126 TCATCTGGTTTCATCGGAAAGTTCGAAAGAGTGATGATGATGATGATGATGATGAT 185  
 Db 980 TCGGTTCTTTTCATTTGCAATAGTTCAAACTTTCGAAACCAATGTCAATCTCTTGTCT 921  
 QY 186 TTCCTCGTGGTGGTTCCTATGATGACCTCAAGAGTCCTCCAAATATCAAAAGCCGAA 245  
 Db 920 TTTAGCTGATGGTGGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 861  
 QY 246 TCACAAATTTGAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305  
 Db 860 TTGCATTTGTTGCTATCTATCATATTTCTCCCTGCTCAATGCCATGTTAGAAATAGTTA 801  
 QY 306 GCCCTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATTC 352  
 Db 800 GCTTTAAATTTAATACATGCGCTTACCTTCTTCTGTCACCATC 754  
 RESULT 9  
 AC124959/c  
 LOCUS AC124959 125883 bp DNA linear PLN 05-MAR-2003  
 DEFINITION Medicago truncatula clone mth2-23111, complete sequence.  
 ACCESSION AC124959  
 VERSION AC124959.11 GI:28850030  
 KEYWORDS HTG  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 1 (bases 1 to 125883)  
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 30237: contig of 30237 bp in length  
 \* 30238 30337: gap of unknown length  
 \* 30338 109284: contig of 78947 bp in length.  
 FEATURES  
 source  
 1. 109284  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-59121"  
 /clone\_lib="Medicago truncatula BAC library H2"  
 ORIGIN  
 Query Match 3.9%; Score 83.8; DB 2; Length 109284;  
 Best Local Similarity 54.5%; Pred. NO. 4.7e-10;  
 Matches 189; Conservative 0; Mismatches 157; Indels 1; Gaps 1;  
 QY 7 CAACCTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAAGTTAGAAATCTTTA 66  
 Db 1100 CAAGATCTGGAGACATCTTAAATTTTCTTACTCTTCATGATGTAATCTCTT 1041  
 QY 67 CCAAGAGCTTTGAGTCCATGATGACATCCGTGAAACGGTG-TACATGTCCTCCGATGCAC 125  
 Db 1040 TCAAGAGAGCGTAATCATTAATGATAGTAAACCTTCCATACATTTTCATCTATGTT 981  
 QY 126 TCATCTGGTTTCATCGGAAAGTTCGAAAGAGTGATGATGATGATGATGATGATGAT 185  
 Db 980 TCGGTTCTTTTCATTTGCAATAGTTCAAACTTTCGAAACCAATGTCAATCTCTTGTCT 921  
 QY 186 TTCCTCGTGGTGGTTCCTATGATGACCTCAAGAGTCCTCCAAATATCAAAAGCCGAA 245  
 Db 920 TTTAGCTGATGGTGGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 861  
 QY 246 TCACAAATTTGAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305  
 Db 860 TTGCATTTGTTGCTATCTATCATATTTCTCCCTGCTCAATGCCATGTTAGAAATAGTTA 801  
 QY 306 GCCCTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATTC 352  
 Db 800 GCTTTAAATTTAATACATGCGCTTACCTTCTTCTGTCACCATC 754

# REFERENCE AUTHORS TITLE JOURNAL

2 (bases 1 to 125883)  
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 Direct Submission  
 Submitted (20-JUN-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

# REFERENCE AUTHORS TITLE JOURNAL

3 (bases 1 to 125883)  
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 Direct Submission  
 Submitted (06-NOV-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

# REFERENCE AUTHORS TITLE JOURNAL

4 (bases 1 to 125883)  
 Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 Direct Submission  
 Submitted (05-MAR-2003) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

# COMMENT

On Mar 5, 2003 this sequence version replaced gi:24635962.  
 ----- Genome Center  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR

# FEATURES source

Location/Qualifiers  
 1. 125883  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-23111"  
 /clone\_lib="Medicago truncatula BAC library H2"

# ORIGIN

Query Match 3.9%; Score 83.6; DB 8; Length 125883;  
 Best Local Similarity 51.4%; Pred. No. 5.3e-10;  
 Matches 257; Conservative 0; Mismatches 224; Indels 19; Gaps 2;  
 QY 1 GGATCCCAACCTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTTCAAAAGTTAGAAAA 60  
 Db 5238 GGTCTCCATACACCTTGAAGACATCTTAATATTTTCTGATTTCTATCATAGTAGAATA 5179  
 QY 61 TCCTTACCAAGAGCTTTGAGTCCATTCATGATGATCCGTGAAACG-GTGTACATCTCTCG 119  
 Db 5178 GCCTTTTCCAGAGATCTCATTTCTACTTATTGTTAGTAAATCTAGCATACATCTCATCA 5119  
 QY 120 ATGGACTCAGTGGTTTCATTTCGAAAAGTTCGAAAGAGTGCATAAGAAATATTGATTTTG 179  
 Db 5118 ATGGTTTCATTTTCACCTCATTTTCAAAACTTCAAAATTTCTGACTCCAATATCAATCTTA 5059  
 QY 180 GATTTCTTCACTCGGTTGGTGCCTTCATGATGATGATGATGATGATGATGATGATGATGAT 239  
 Db 5058 GTTCTTTGACATGAGATGTACCTCATGATATATTTTCAGAGTGTCCCATACCTCTTTG 4999  
 QY 240 GCCCAATCAAAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299  
 Db 4998 GCATTTTACATCATCACTCTTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4939  
 QY 300 TTCTAGCTTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTCCTCATC 359  
 Db 4938 AGTCTAGCTTTAGAGTTTAGTAATACATGAGCCCTTTTCATCTGTGTGACCATGCTTCTTT 4879  
 QY 360 GGAAGAGAAATTTT-----GAAATCCATTTTTCGCAATGACCAAA 401  
 Db 4878 TCCTTTATAGTCTCTTCTTTTGTGTGGAACAAAGTCTCCATCTGTGATACACTGACCAAC 4819  
 QY 402 AGCTCGAAATCCATGGAATCGAAGACATCTCATATGATGTTTCCCATATCATGATTTTC 461  
 Db 4818 ATATCATTTCTTGGATCTGAGAAAGAGATTTCCATCTTACCTTTCCAGAGTAGTAGTTT 4759  
 QY 462 GACTCATTAACATAGGTGG 481



\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

* 1 4380: contig of 4380 bp in length
* 4381 4480: gap of unknown length
* 4481 14258: contig of 9778 bp in length
* 14259 14358: gap of unknown length
* 14359 20669: contig of 6311 bp in length
* 20670 25083: gap of unknown length
* 25084 25183: gap of unknown length
* 25184 41406: contig of 16223 bp in length
* 41407 41507: gap of unknown length
* 41508 60351: contig of 18845 bp in length
* 60352 60451: gap of unknown length
* 60452 87167: contig of 26716 bp in length
* 87168 87267: gap of unknown length
* 87268 141590: contig of 54323 bp in length.
FEATURES
    source
    1..141590
        /organism="Medicago truncatula"
        /mol_type="genomic DNA"
        /db_xref="taxon:3880"
        /clone="mth2-33121"
        /clone_lib="Medicago truncatula BAC library H2"
  
```

## ORIGIN

```

Query Match      3.8%; Score 83; DB 2; Length 141590;
Best Local Similarity 54.1%; Pred. No. 7.5e-10;
Matches 190; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

Qy 1 GCATCCCAACTTTAGGAATGGATCTTAATAATTTTAGTTATAGTTCAAGTTAGAAAA 60
Db 40054 GGCTCCATGGCTTGAAGACATCTTAAGATTTTCTTATCCGATCATGATGGTATA 40113
Qy 61 TCTTTACCAAGAGCTTTGAGTCCATTTGATGATCATCCGTGAAAC-GGTGTACATGCTCCG 119
Db 40114 GTTTTTCAGTGAATCTTAATCAATCATACATAGATGTGATCTCGGAGTACATTTATCA 40173
Qy 120 ATGACATCATTGGTTTCATTCGAAAGTTTGAAGAGAGTGATCAATGAATATTCATTTTG 179
Db 40174 ATGTTCTCATCTCGTTTCATTTCAAAGATTTCAAATTTCTTACGCCAAATGCTGATTTCT 40233
Qy 180 GATTCTTTTCACTCGTTGGTGCCTTCATGATGACCTCAGAGTCCTCCAAATATCAAAA 239
Db 40234 GTTTCTTTTACATGGCTAGTTCCTTCATGATGAATTTTATAGAGTGTCCTCCACACTTCTTT 40293
Qy 240 GCCGAATCACAAAATGAAATGTGATTCGAATTCATTTTGTCTAATGCACAAAACAGGCA 299
Db 40294 GCAGTCACACATTCATCACTCTTTCATCTTCTTCCATGCTTAGTGGCATGATGAAG 40353
Qy 300 TTCATAGCTTTGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCAT 350
Db 40354 AGTCGAGCTTTAGAGTTTAGAAGTACTTCTCTTTATCTTCTTTTGACCAT 40404
  
```

## RESULT 12

```

AF411804
LOCUS              100810 bp DNA linear PLN 25-JUL-2002
DEFINITION        Lycopersicon esculentum BAC clone Clemson_ID 47113, complete
sequence.
ACCESSION         AF411804
VERSION           AF411804.1 GI:15987769
KEYWORDS
SOURCE
ORGANISM
  Lycopersicon esculentum (tomato)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
  1 (bases 1 to 100810)
  Van Der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and
  
```

## TITLE

Tanksley, S.  
 Deductions about the Number, Organization, and Evolution of Genes  
 in the Tomato Genome Based on Analysis of a Large Expressed  
 Sequence Tag Collection and Selective Genomic Sequencing

## JOURNAL

Plant Cell 14 (7), 1441-1456 (2002)

## PUBMED

12119366

## REFERENCE

2 (bases 1 to 100810)

## AUTHORS

van der Hoeven, R.S. and Tanksley, S.D.

## TITLE

Direct Submission

## JOURNAL

Submitted (20-AUG-2001) Plant Breeding, Cornell University, 252  
 Emerson Hall, Ithaca, NY 14850, USA

## FEATURES

Location/Qualifiers

1..100810

/organism="Lycopersicon esculentum"

/mol\_type="genomic DNA"

/db\_xref="taxon:4081"

/clone="Clemson\_ID 47113 BAC"

## ORIGIN

```

Query Match      3.8%; Score 81.2; DB 8; Length 100810;
Best Local Similarity 52.0%; Pred. No. 2.1e-09;
Matches 205; Conservative 0; Mismatches 188; Indels 1; Gaps 1;

Qy 4 TCCCACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAAGTTAGAAAAATCT 63
Db 89844 TCCCAAGACCTAGAAAAATCCTAAGTACTTTGTCAACCTGTTCTCAGTGGTATATACT 89903
Qy 64 TTACCAGAGCTTTGAGTCCATTTGATGACATCCGTGAAAC-GGTGTACATGTCCTCCGATG 122
Db 89904 TTACCAATGAGATTTAGTTTTCGTTTATCACAGTGTGTAATCTAGTAATCATATCTGA 89963
Qy 123 GACTCATTGGTTTCATTCGAAAGTTTCGAAAGAGTGATCAATTAATTTGATTTTGGAT 182
Db 89964 GATTTCCGATTTCACTTTAAGGCTTCATTTCTGACACACAACATAGCAATCTTGAAT 90023
Qy 183 TCTTTCACTCGTTGGTGCCTTCATGAGTCACTCAAGAGTCCTCCAAATATCAAAAGCC 242
Db 90024 TTTCCGACTTTGAGTTGACCTTCATGAGAGTTTGTAAAGTATCCCATATTTGTTTAGCA 90083
Qy 243 GAATCAAAATGAAATGTGATGAATTCATTTTGTCTAATGCACAAAACAGGCGATTC 302
Db 90084 GAGTACAACCTGAGATACGTTTAACTCATCGGTCCTAGTCCACATCAAGAATGCAC 90143
Qy 303 ATAGCCTTTGTTTAAAGCAAAAACATTCCTCCGATTCATCCCATTCGCTCATCGGA 362
Db 90144 TTAGCTTTGCATCTTCTCCATGATTTTATATCCGCTTCATATATGATCTGCGAGT 90203
Qy 363 AGAGAAATTTTGAATCCATTTTCGACAATAG 396
Db 90204 TTAGGTACTTTATTTCTCTCACTATCAGTAATAG 90237
  
```

## RESULT 13

```

AC087797
LOCUS              122497 bp DNA linear PLN 29-AUG-2001
DEFINITION        Oryza sativa chromosome 3 BAC OSUNB0022E02 genomic sequence,
complete sequence.
ACCESSION         AC087797
VERSION           AC087797.5 GI:13699786
KEYWORDS
SOURCE
ORGANISM
  Oryza sativa (japonica cultivar-group)
  Oryza sativa (japonica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
  1 (bases 1 to 122497)
  Ruell, C.R., Yuan, Q., Ouyang, S., Moffat, K.S., Hill, J.N.,
  Gansberger, K., Brenner, M., Burgess, S., Hance, M., Shvartsbeyn, M.,
  Tsitin, T., Riggs, F., Hsiao, J., Zismann, V., Blunt, S., Pai, G.,
  VanAken, S.E., Utterback, T.R., Feldblyum, T.V., Quackenbush, J.,
  Salzberg, S.L., White, O. and Fraser, C.M.
  Oryza sativa chromosome 3 BAC OSUNB0022E02 genomic sequence
  Unpublished
  
```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (25-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
3 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (03-MAY-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuella@tigr.org
5 (bases 1 to 122497)
Buell,R.
Direct Submission
Submitted (29-AUG-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, rbuella@tigr.org
On Apr 20, 2001 this sequence version replaced gi:13603985.
Address all correspondence to:rice@tigr.org

BAC clone OSJNB0022E02 is from Oryza sativa chromosome 3
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Egenesh (http://www.softberry.com/),
Genscan and Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), and GeneSplicer
(Mihaela Pertea and Steven Salzberg, contact.mperea@tigr.org),
searches of the complete sequence against a peptide database and
the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as unknown proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as hypothetical proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES
source
1. .122497
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="3"
/map="C944"
/clone="OSJNB0022E02"
/complement(1271..1579)
/gene="OSJNB0022E02.11"
/notes="predicted by fgenesh"
/complement(1271..>1579)
/gene="OSJNB0022E02.11"
/complement(1271..1579)
/gene="OSJNB0022E02.11"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK50135.1"
/db_xref="GI:13937304"
/translation="MENTKAAPLVSSPVSLSLPAADPRPTLRAPLLHPGLSVAP
SCRTRSSGTLTEVPVTKLMAKVVVDDGWSHSVAFQKWAFLAENLAGSVFAATS
"
1386..1640
/gene="OSJNB0022E02.18"
/notes="predicted by glimmer"
<1386..>1640

/feature="OSJNB0022E02.18"
1386..1640
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK50133.1"
/db_xref="GI:13937302"
/translation="MRASVFTGVHNSVPLRVLQATLRPGSSGARNVGRGSTAA
GSDTNGEDDINGAAVFSTAPAQPNQTTGATVREEL"
complement(5240..12557)
/gene="OSJNB0022E02.2"
/notes="EST D47656, AU088605 from this gene"
complement(join(5240..5517,5605..5684,6132..6212,
6314..6383,6539..6613,6693..6898,7177..7426,7506..7551,
8773..8842,9564..9633,11257..11341,11440..11524,
11608..11698,12255..>12557))
/gene="OSJNB0022E02.2"
complement(join(5439..5517,5605..5684,6132..6212,
6314..6383,6539..6613,6693..6898,7177..7426,7506..7551,
8773..8842,9564..9633,11257..11341,11440..11524,
11608..11698,12255..12388))
/gene="OSJNB0022E02.2"
/codon_start=1
/product="unknown protein"
/protein_id="AAK50136.1"
/db_xref="GI:13937305"
/translation="MORRAQTWAGVGTQAQAAHAALFCFTLLALKVDGRTATYSW
WIIFPLWFGHIVAGRFSPAPSLPHGRHAPCHSIVAAPLIIAFELLCILLES
RYKSPVTDLKIVFLPLFAFEVILADNFRMCALMGDESDAIEWETLPHFWVA
ISWFLIATTTLLKSGDVGALGWDLFINYIAECFAFLVCTRWFPNTHKSPNP
GEASSSAAIRYRDWESGLLPSLEDHEOERLCGLPDIGGHVWKPLVIFVOLLWML
ECTPSPAQYIPFALFSPFLIQAGVLSLARLLEKLVLLRNGDVSNLYLTISKV
RCCFALHRSGLGWSIDEGKEEQARLFYTESTGYNTFCGYPEVVRKMPKRDLA
RQVRLQAALGEQSEITCKTQCFERLQNEKVLRCICEYEICMVLPLCRHRTLCTC
SDKCKKCPICRVPIERMEVYDV"
complement(12328..12420)
/rpt_family="(CGGGGG)n"
/rpt_family="(GA)n"
complement(12558..12580)
/rpt_family="(GA)n"
14168..18366
/gene="OSJNB0022E02.12"
/notes="similar to gag pol polyprotein GB:AAD12997
GI:4234852 (Zea mays)"
join(<14168..15023,15131..15274,15344..16687,16818..17803,
17887..18048,18262..>18366)
/gene="OSJNB0022E02.12"
join(14168..15023,15131..15274,15344..16687,16818..17803,
17887..18048,18262..18366)
/gene="OSJNB0022E02.12"
/codon_start=1
/product="putative gag-pol polyprotein"
/protein_id="AAK50117.1"
/db_xref="GI:13937286"
/translation="MAGFADALRPDKFTGVHFRKWIRVTLMTAMKCFWSTGKPEG
VLTAQKQKFEATTLFVGCILSLVGLDRIVEVYHMTDAKELWDLNTFEGATDASND
LYIMEQFDHYKMDNRSVVEQAEHETQWAKELLCVLPDKFVAGCIITAKLPSPWRS
FGTALKHRRQYSVEGLIASLDEKAREKDAASKDGQGSANVHKAQSKGKYK
AQQTTFKKKNNNNPNQDTCFVQGVGHARKQPKQKMGKAPAGTQSKSANVTI
GNTDGSYGRTRGPHNGEWTFCSCWCHGRSEVYFGKRAAERAAACPFYRQESC
YECGLFRSLSDFCNKHCAEENLAPLELLSDCLCMNGVLTGKGRYPMLID
IVKSKCHSCVQSKPRKPKHCAEENLAPLELLSDCLCMNGVLTGKGRYPMLID
DATRCYVLLTKDEALDYFKYAEVENQIDRKIKRLSRDGEFFSNFEDLPCEE
HGIHERTPFYPSPESNGIAERKNRTLTDLVNMALDTAGLPAKWGEALLTSNVLNRV
PNRNKDKTPYEIWIQRKPSLYRTWGLAKVNPITKRKLGPRTVDCVFLGVAHHS
LAYRELIYKSEVPMHVGTIMESEADTEFFEFPMKTHSGNSQSEIIPSSITPEEQ
TEHTHELVSDEVSAPRRSKORTAKSGDDFTVLYVDDTPTKSEAVASPDADYWK
BAVRSEMDIILANGTWETPSPYCKPVCKVKKLRLPADASHGLLHQMDVKTAFL
NGELDEEYIMQPDGFFVVEGEGKCKLLSLYGLQKAPQKHEFKDTLTSAGAVN
EADCKVYRHGGEGVILCLYVDDILIFGTNLNEVINEVKSFLSQFMDKDLGAVIL
NIKIRGENGITLLOSHYVEKILNRFTSNPDGDHWRALERVMRYLKGVEGLHTVG
QIIGSLMYLAGATRPDI3FAVSKLSRFTSNPDGDHWRALERVMRYLKGVEGLHTVG
YPAVLEGYSDSNWISDVEIKATSGVFTLGGAVSWRSCKQTLITRSTMEALTALD
TATVAEAWLRDLLMDLPVVEKPTCEKTIETVQCEIKLRSYNVGLHFNSEKPSRSHGEG

```



```
TTTKDQCQIEGPGFETHSILEGTHLCELDWCWSQSMKILGESSRKLTKDLGV"
complement(18822..18858)
/rpt_family="CAGGC)n"
repeat_region
complement(19612..19640)
/rpt_family="AT rich"
repeat_region
complement(20459..20487)
/rpt_family="AT rich"
repeat_region
21271..21576
/gene="OSUNBB0022E02.13"
/notes="predicted by fgenesh"
mRNA
<21271..>21576
/gene="OSUNBB0022E02.13"
CDS
21271..21576
/gene="OSUNBB0022E02.13"
/codon_start=1
/product="hypotheical protein"
/protein_id="AAK50120.1"
/db_xref="GI:13937289"
/translation="MTTRTVLLSEYSDSHNSGRNGAVGVGHGVGAGGRGLHGYI
ACAVVLLASTPEVAPLHTQGELOHGVWTKDCEWRQCAPAWDVDEFTSNDASVPNS"
complement(22507..22529)
/rpt_family="TA)n"
repeat_region
complement(22544..22588)
/rpt_family="AT rich"
repeat_region
complement(23124..23217)
/rpt_family="(GGAGAA)n"
repeat_region
complement(23165..25551)
/gene="OSUNBB0022E02.14"
/notes="predicted by fgenesh"
mRNA
complement(join(23165..23588,23765..23917,25379..>25551))
/gene="OSUNBB0022E02.14"
CDS
complement(join(23165..23588,23765..23917,25379..25551))
/gene="OSUNBB0022E02.14"
/codon_start=1
/product="hypotheical protein"

Query Match      3.8%; Score 81; DB 8; Length 122497;
Best Local Similarity 48.5%; Pred. No. 2.4e-09;
Matches 314; Conservative 1; Mismatches 321; Indels 11; Gaps 3;

QY 4 TCCCAACTTTTGGAGTGGATCTTTAAATTTTACGTTTAAAGTTCAAAAGTTAGAAAATCT 63
Db TCATATCTTCGGGTAGGCTCTAAGCATCTTTTGGGCAACCTCAAGATCGGTAGTTT 62967
QY 64 TTACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAAC-GGTGTACATCTCCCATG 122
Db GCCCAAGCCCTTGAGATCATTTTACAAATGACATGTAGCCCTCCCATATCATATTCACA 63027
QY 123 GACTCACTTGGTTTCATTCGAAAGTTGCAAGAGTGCAATAAGATATTCATTTGGAT 182
Db CTCTCATGAGCAACATGGAGATGTCTCATATTGGATCTTGAGAAATGAGCTTGGCA 63087
QY 183 TCTTTCATCGGTTGGTCCCTTCATGATGACCTCAAGAGTCCTCCAAATATCAAAAGCC 242
Db TCCCTGTACTCAC'TGTATCCCTCATGATCTCCGCAACTTGTTCMAATCTCATATGCG 63147
QY 243 GAATCACAATTTGAAATGTGATTAATTCATTTTGTCTAATGACAAAGAGGCATTC 302
Db GTCTCAAGGTTGCTCCTCATCTATCGAATCTCTTGGCTCAAGAGTTGAAACAGGCATTC 63207
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATTCCTCATCGGA 362
Db ATGGC-----TTGGGCATTGATTTGGAGTTGGGTTGGGTTGATCAATCTCCGTCAAGGT 63258
QY 363 AGAGAAATTTTGAATTCATTTTTCGACATAGACCAAGCTCGAATTCATTCGAAATG 422
Db GTGCCGGTGTAGGCAAGCCCTACATCCCAATATCTCCAAATATGGAAGCTCATAGCTTTG 63318
QY 423 AGGAAGATCCTCATATGATTTTCCAAATACATGTAATTCGACTCATTAACATAGGTGGA 482
Db AGGTGAGTAGACATTTTAAATTTTCCAGTGGAGTAGTTTGGCCATTAACATGGAGCC 63378
QY 483 TGTGTAATGAATGACCCCTCATGCSCTATC-TCTCTTTGGGTATTAAACCAAAATATGAG 541
```

```
Db 63379 TTCCCTACATGGTTACCTTCGTCATCTTCTCTAGCGGTGAAGCCCAATCAAGA 63438
QY 542 TGAGCCCTTGTCTGTATACCAATTTGTTAGATCAGATGGCACTAAGAGAGGGGGAGTG 601
Db 63439 GAGACCAAGCTCTGATACCACTTTGTAGGATCGAGATGTCGACTAGAGGGGGTGAATAGG 63498
QY 602 AATTAGTCGAGTGGATTAAACCTTATAAGTTTAAAAATGAATTCGTA 648
Db 63499 CGATTTAAACTAAATAACACCTAATCAAACTAACCTAAGTTGCTA 63545
```

## RESULT 14

AC144740 157436 bp DNA linear HTG 15-MAY-2003  
LOCUS Oryza sativa (japonica cultivar-group) chromosome 5 clone  
DEFINITION OSUNBa0075A10, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 9 ordered pieces.  
ACCESSION AC144740  
VERSION AC144740.1 GI:30725895  
KEYWORDS HTG; HTGS PHASE2.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 157436)  
AUTHORS Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,  
Chao,Y.-T., Lee,P.-F., Chang,S.-J., Chen,H.-C., Chen,S.-K.,  
Hsiao,S.-H., Hsiung,J.-N., Hsu,C.-H., Kau,P.-I., Lee,M.-C.,  
Liu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,  
Yu,S.-W., Wu,H.-P. and Shaw,J.-F.

Oryza sativa BAC OSUNBa0075A10 genomic sequence

## TITLE

Unpublished  
2 (bases 1 to 157436)  
Chow,T.-Y. and Hsing,Y.-I.C.

## REFERENCE

Submitted (15-MAY-2003) Institute of Botany, Academia Sinica, 128,  
Section 2, Academia Road, Nankang, Taipei 11529, Taiwan

## AUTHORS

Direct Submission

## JOURNAL

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

## COMMENT

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 5300: contig of 5300 bp in length  
\* 5301 5400: gap of unknown length  
\* 5401 17970: contig of 12570 bp in length  
\* 17971 18070: gap of unknown length  
\* 18071 37874: contig of 19804 bp in length  
\* 37875 37975: gap of unknown length  
\* 37976 62822: contig of 24848 bp in length  
\* 62823 62923: gap of unknown length  
\* 62924 64710: contig of 1788 bp in length  
\* 64711 71025: gap of unknown length  
\* 71026 71126: contig of 6215 bp in length  
\* 71127 118711: contig of 47586 bp in length  
\* 118712 121761: gap of unknown length  
\* 118812 121761: contig of 2950 bp in length  
\* 121762 121861: gap of unknown length  
\* 121862 157436: contig of 35575 bp in length.

## FEATURES

## source

1..157436  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="5"  
/clone="OSUNBa0075A10"



## ORIGIN

Query Match 3.8%; Score 81; DB 2; Length 157436;  
Best Local Similarity 48.5%; Pred. No. 2.4e-09;  
Matches 314; Conservative 1; Mismatches 321; Indels 11; Gaps 3;

QY 4 TCCCACTTTTGGAAATGGATCTTAAATTTTGGTTATTAAGTTCAAAAGTTAGAAAAATCT 63  
Db 96233 TCATACCTTCTCCGGTAGAGGCTAAGCATCTTTTGGGCAACTCAAGATCATGTAGTTT 96292

QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCCCGTGAAC-GGTGTACATCTCCCGATG 122  
Db 96293 GCCCAAGCCCTTTGAGATCATTTACATGATCAATGAGCTCCCATATCATGCTATTACA 96352

QY 123 GACTCACTTGGTTTCATTCGGAAGTTCGAAAGATGTCATAAGAAATATTGATTTGGAT 182  
Db 96353 CTCTCATGAGCAACATGGAGAATGCTCATATTGGATCTTTAGGAATGAAGCTTGCCA 96412

QY 183 TCTTTCACCTGGTTGGTCCCTTCATGATGACACCTCAAGAGTCCCTCCAAATATCAAAAGCC 242  
Db 96413 TCCTTGTACTCGTTTGGCCCTCATGGATCTCCGCAACTTGTCCAAATCTCATATGCG 96472

QY 243 GAATCACAAAATTGAATGTGATTTGAATTCATTTTGTCTAATGACACAAACAGGCGCATTC 302  
Db 96473 GTCTTAAGTTGTCTCACTCTATCGAACTCTCTTGGCTCAAAGAGTTGAACAAAGCATTC 96532

QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATCTCTTCCGATTCATCCATTCGCTCATCGGA 362  
Db 96533 ATAGCTTGGCATTTGATTTGGAGTTGTGGTGTGATCAATCTCCGTCAAAGGGGCTCGGTG 96592

QY 363 AGAGAAAATTTTGAATCCATTTTCGACATATGACCAAGCTCGAAATCCATGGAATG 422  
Db 96593 ATAGTAAAGCCTACA-----TCCACATATCTCCAATATGGAACCTCATAGCTTTG 96643

QY 423 AGGAAGTCTCTATGAGTTTCCAAATACATGTAATTCGACTCATTAACATAGGTGGA 482  
Db 96644 AGGTGATGACATATTTAAATTTTCCAAAGTGGAGTGTGTGCTTCCATTAACATGGGAGCC 96703

QY 483 TGTGTAATGAATGACCTCATGCTATC-TCTCTGGGTATTAACCAATATGAGAG 541  
Db 96704 TTCCTTACATGGTTTCACTCGTTCGACATCTCTCTAGGGCGGTAAAGCCCAATCAAGA 96763

QY 542 TGAGCCTTGTCTGATACCAATTTGTTAGGATCAGATGGCACTAAGAGAGGGGGGAGTG 601  
Db 96764 GAGACCAAGCTCTGATACCACTTGTAGATCGAGATCTCACTAGAGGGGGGTGAATAGG 96823

QY 602 AATTAGTGCAGTGGATTAACATTAAGTTTAAATGAATTCGTA 648  
Db 96824 CGATTTAAATAAATGACACCTAATCAAAACTAATCACTAAGTTACTA 96870

RESULT 15  
AP003198/c  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,  
BAC clone: B1039D07.  
AP003198 BA000010  
VERSION  
AP003198.3 GI:15408621  
KEYWORDS  
SOURCE  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE  
AUTHORS  
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,  
Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nakamura, Y.,  
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, Y.,  
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,  
Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, Y.,  
Itonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Itoh, S., Itoh, T.,  
Itoh, Y., Itoh, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,  
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,

Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,  
Nakama, Y., Nakamichi, Y., Nakamura, M., Mamiki, N., Negishi, M.,  
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,  
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Tsuji, K., Waki, K.,  
Yamagata, H., Yanane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,  
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,  
Yano, M., Jiang, J. and Gojobori, T.  
The genome sequence and structure of rice chromosome 1  
Nature 420 (6913), 312-316 (2002)  
22337376  
12447438  
2 (bases 1 to 135301)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (19-FEB-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
On Aug 31, 2001 this sequence version replaced gi:14522858.  
Genes were predicted from the integrated results of the following:  
GENSCAN1.0, BLAST2.0, BLASTX2.0 as well as SplicePredictor  
(October 1998 version). The genomic sequence was searched against  
NCBI NonRedundant Protein database, nr  
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at  
RGP. Protein homologies of the coding regions were searched against  
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent  
the identified cDNA sequences using BLASTN 2.0 with the  
corresponding DBJ accession no. and RGP clone ID.  
A gene with identity or significant homology to a protein is  
classified based on the protein name to indicate the homology level  
such as same name, 'putative-' and '-like protein'. A gene without  
significant homology to any protein but with EST homology (covering  
almost the entire length of partial sequence) is classified as an  
'unknown' protein. A gene predicted with a gene prediction program  
is classified as a 'hypothetical' protein.  
The orientation of the sequence is from M13rev to -21M13 of the BAC  
clone. This sequence of B1039D07 clone has an overlap with P0493G01  
(DBJ: AP002914) clone at the position 1 to 55,572 and with  
B1045D11 (DBJ: AP003199) at the position 63,818 to 135,301 of 5'  
end. The sequence of this clone starts at the position 11,2015 of  
P0493G01 and ends at the position 71,483 of B1045D11. Detailed  
information on overlap and assembly quality together with  
annotation of this entry is available at  
http://rgp.dna.affrc.go.jp/GenomeSeq.html.  
FEATURES  
Location/Qualifiers  
1..135301  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="1"  
/clone="B1039D07"  
complement(join(1884..1995,2143..2186))  
/gene="B1039D07.1"  
complement(join(1884..1995,2143..2186))  
/gene="B1039D07.1"  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAB64039.1"  
/db\_xref="GI:15408622"  
/translation="MTAGGEEGEGVRSSWFTTSGLIYEEVVTTRDMTAGGEEERGE  
GYRQWD"  
complement(join(4700..5160,5990..6038,6142..6291))  
/gene="B1039D07.2"  
complement(join(4700..5160,5990..6038,6142..6291))  
/gene="B1039D07.2"  
/codon\_start=1  
/product="putative glutathione S-transferase"  
/protein\_id="BAB64040.1"  
/db\_xref="GI:15408623"  
/translation="WTPKVFQPAQSTNVARVLLCLLEVGAEYEVNVDFVTMEHKSP  
EHLKRNPFQIPAFQDGLDLYFESRAIGKYILRKYTKREADLLREGNREAAVVDVMT

```

EVTHOYNSAISPIVVECIINPAMRGITPTNQKVVDRAEKLKKVLEVEARLSOSTVL
AGDFVSFADLNHFPIFYFMGTYPASLFDSYPHVKAWERLMARPSVKYKLAAYWAPQG
A"
join(12552..12701,13864..13940,14245..14653)
/gene="B1039D07.3"
CDS
join(12552..12701,13864..13940,14245..14653)
/gene="B1039D07.3"
/codon_start=1
/product="putative glutathione S-transferase"
/protein_id="BAB64041.1"
/db_xref="GI:15408624"
/translacion="MAPVKVGPAPKSTAVARVLVCLBEVGAIEYELVGIHIPAGEOKSP
AHLARNDESKCVNFDLILKSSKHKTCRVFCASDLILKESNLSSAMVDVWLEVESQ
TFDTAMSAITFOCLITFTPGGIADDAKIVENLGLKKALEYEARSCRYLAGDFI
SLADLSHPMTHYLILATPHASVLDAYPHVKSWINDLMKRFAVKRVREIMEA"
join(15075..15224,16017..16065,17096..17547)
/gene="B1039D07.4"
CDS
join(15075..15224,16017..16065,17096..17547)
/gene="B1039D07.4"
/codon_start=1
/product="putative glutathione S-transferase"
/protein_id="BAB64042.1"
/db_xref="GI:15408625"
/translacion="MAPAKVYGPAMSTNVMRILVCLBEVGAIEYEWVPMVMSTGEHKRP
PHISENPGQVPAEDGDLITFEISRAISKVILRKHGSDLLRESNLSESAMVDVWLEVE
SSHFDGAMSPFIQCFIVPMGATIGVNVESLEKLKALEYVEAQLSKSVLAGD
PISLADISHFTPTVYLLASAHVLEAYPRKVIADDDVMQRPVKKVTEALKMPSA"
join(19985..22027,46611..47316)
/gene="B1039D07.5"
CDS
join(19985..22027,46611..47316)
/gene="B1039D07.5"
/misc_feature
/note="probably inactive due to frameshift(s) in CDS
probably inactive due to no initiation/termination codons
in CDS
probably inactive due to stop codon(s) in CDS
pseudogene
similar to polyprotein"
LTR
complement(22028..22596)
/note="3' LTR"
gene
join(23115..23315,23412..23720)
/gene="B1039D07.6"
CDS
join(23115..23315,23412..23720)
/gene="B1039D07.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64043.1"
/db_xref="GI:15408626"
/translacion="MAPRKNPASATGPDGRIDDSTAFLGASLVDDDELAKLVSAQ
PLVEQAEPLGKAVPKGDNTWLSLSALVRVALFDWACRTAGPEPSTELFDVVFYAT
VNSKTVVTPARTKKTAFGVSFNVNRSRSLRPNVNAAMSKDRHMAKWFYHSIPFET
GSDAAKALR"
join(24175..24595,31053..33108)
/gene="B1039D07.7"
misc_feature
join(24175..24595,31053..33108)
/note="probably inactive due to frameshift(s) in CDS
probably inactive due to no initiation/termination codons
in CDS
probably inactive due to stop codon(s) in CDS
pseudogene
similar to non-LTR retroelement reverse transcriptase"
LTR
24597..25567
/gene="B1039D07.7"
/note="5' LTR"
gene
25604..29509
/gene="B1039D07.8"
misc_feature
25604..29509
/gene="B1039D07.8"
/note="probably inactive due to stop codon(s) in CDS
pseudogene
similar to retrovirus-related pol polyprotein from

```

```

transposon TNT 1-94"
/pseudo
30069..31051
/gene="B1039D07.7"
/note="3' LTR"
join(35207..35356,35417..35752)
/gene="B1039D07.9"
join(35207..35356,35417..35752)
/gene="B1039D07.9"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64044.1"
/db_xref="GI:15408627"
/translacion="MEALCLEVSKAEASAEKESORFVREMAKATESAMACQTLRLAL
SDMEARAGGAVSDCATYAGDCCARVSAFTMGFLOPGFYEHAIEFPYAKGDMVEVSAQ
DISPALHAWRKQIWKQDGRSTAKARLLEQLAKAARADQGEEDAAVGGGGDAQDHPE
V"
complement(join(36671..36837,37475..37856))
/gene="B1039D07.10"
complement(join(36671..36837,37475..37856))
/gene="B1039D07.10"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB64045.1"
/db_xref="GI:15408628"
/translacion="MVAAGDHRGGAAAEERGEKGEKRRSTAHPRSTATTKKAAGAE
EGGAARVDGVDGVPVAGERNGGVDEVEDAAKPEAAPRWVVRGDDGGPELNGD
GERRRRELDSGEDGRGAETDGGVILGRDRNENSLFRKTNRTPLHSINMKTAHL
SAQNTTDRKRKSKAKKSADLFR"
complement(40236..45213)
/gene="B1039D07.11"
complement(40236..45213)
/gene="B1039D07.11"
/note="probably inactive due to frameshift(s) in CDS
Query Match 3.7%; Score 80.6; DB 8; Length 135301;
Best Local Similarity 49.4%; Pred. No. 3e-09;
Matches 295; Conservative 1; Mismatches 290; Indels 11; Gaps 3;
QY 4 TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATATAGTTCAAAGCTTAGAAAAATCT 63
DB 95099 TCATACTTCTCCGGTAGGGCTTAAGCATCTTTGGGCAACCTCAAGATCGGTGATGTT 95040
QY 64 TTACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAAC-GGTGTACATGCTCTCCGATG 122
DB 95039 GCCCAAGCCCTTGAGATCAITCACAATGACATTAACCTCCCATACATGTCATCACA 94980
QY 123 GACTCACTTGGTTCAITTCGGAAAAGTTCCGAAAGAGTCGATAGCAATATGATTTGGAT 182
DB 94979 CTCTCATGAGGCAACATGGAGAATGCTCTCATATTGGATCTTTGAGGAAATGAAGCTTGGTA 94920
QY 183 TCTTTCACTCGTTGGTGGCTTCATGAGTGACCTCAAGAGCTCTCCCAATATCAAAAGCC 242
DB 94919 TCCTTGTACTCACTTGTAACCTCATGGATCTCCGCCAACTTGTTCCAAATTCATATGG 94860
QY 243 GAATCACAATTTGAAATGTGAATTTGATTTTGTCTTAATGTCACAAAACAGGCAATTC 302
DB 94859 GTCTCAAGGTTGCTCACTCTATTGAACTCTCTTGGCTCAAGAGTTGAACAAGCAATTC 94800
QY 303 ATAGCTTTGTGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCATTCGCTCATCGGA 362
DB 94799 ATGGCTTGGCATTTGAGTTGAA-----GGTTGCGGTGATCAATCTCGCTCAAGGT 94749
QY 363 AGAGAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGAAATCCATGGAATG 422
DB 94748 GTGTCGGTAGAGCAAGCCCTACATCCCAATACATCCCAATATGGAAGCTCATAGCTTT 94689
QY 423 AGGAAGATCCTCATATGAGTTTTCCAATACATGTAATTCGACTCATTAACAATAGTGA 482
DB 94688 AGGTGAGTAGACATTTTAATTTTCCAAAGTGGAGTAGTTGTGCCATTAAACATGGAGCC 94629
QY 483 TGTGTAATGAATGACCTCATGCSCTATC-TCTCTTGGGTATTAAACCAATATGAGAG 541

```

Db 94628 TTCCCTACATGGHTTCACTCGTTGACATCTTCTCTCTAGGTGGTGAAGCCCAATCAAGA 94569  
 QY 542 TGAGCCTTGCTCTGATACCAATTGTTAGGATCAGAGTGGCACTAAGAGAGGGGGGGA 598  
 Db 94568 GAGACCAAGCTCTGATACCACTTGTAAAGATCGAGATGTCGACTAGAGGGGGGTGAA 94512

Search completed: September 18, 2004, 23:51:53  
 Job time : 8127.47 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:18:43 ; Search time 761.294 Seconds  
(without alignments)  
12030.986 Million cell updates/sec

Title: US-09-892-635A-44

Perfect score: 2156

Sequence: 1 ggatcccaacttttagaat.....agctctcattcactatgg 2156

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002s:\*

7: geneseqn2003as:\*

8: geneseqn2003bs:\*

9: geneseqn2003cs:\*

10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2120.8	98.4	4924	2	AAX25614 Banana ri
2	75	3.5	809	2	AAX28690 Ripening
3	62.8	2.9	2417	2	AAX22222 SIRE-1 CD
4	54	2.5	110000	3	Continuation (3 of
5	53.4	2.5	110000	3	Continuation (4 of
6	52.6	2.4	515	2	AAG81475 HMG2 prom
7	51.2	2.4	2000	7	ADA71938 Rice gene
8	46.8	2.2	16439	6	ABL32886 Human imm
9	46.2	2.1	2000	7	ADA71938
10	45	2.1	742	2	AAX28672
11	44	2.0	1186	2	AAX25608
12	41.2	1.9	1722	7	ACF73552
13	41.2	1.9	2881	2	AAV74565
14	41.2	1.9	5879	6	ABL32229
15	41.2	1.9	8197	6	ABL34515
16	41.2	1.9	8197	6	ABL70542
17	41	1.9	3825	6	ABX13945 Clostridi
18	41	1.9	3825	6	ABX13945
19	41	1.9	6022	4	AAS42523
20	40.8	1.9	7351	4	AAS46662
21	40.8	1.9	18967	6	ABL32028
22	40.6	1.9	726	6	ABQ68510
23	40.6	1.9	1337	6	ABQ70519 Listeria

## RESULT 1

AAX25614

ID AAX25614 standard; DNA; 4924 BP.

XX

AC AAX25614;

XX

DT 02-AUG-1999 (first entry)

XX

DE Banana ripening fruit chitinase DNA.

XX

KW Banana; fruit ripening; differential expression; fruit development;

KW transgenic plant; chitinase; ss.

XX

OS Musa acuminata.

XX

PN WO9915668-A2.

XX

PD 01-APR-1999.

XX

PF 23-SEP-1998; 98WO-US003343.

XX

PR 25-SEP-1997; 97US-0060062P.

XX

PA (BOYC-) BOYCE THOMPSON INST PLANT RES.

XX

PI May G, Clendennen S;

XX

DR WPI: 1999-244425/20.

XX

DR P-PSDB; AAY05845, AAY05846, AAY05847.

XX

PT New isolated banana DNA molecules.

XX

PS Claim 5; Fig 17A-G; 143pp; English.

XX

CC This is the nucleotide sequence of a chitinase DNA molecule. The invention provides isolated DNA molecules which are differentially expressed during banana fruit development, and the protein products of these genes. The DNA is selected from a group comprising starch synthase, chitinase, endochitinase, beta-1,3-glucanase, thaumatin-like protein, ascorbate peroxidase, metallothionein, lectin and senescence-related protein. The regulatory elements of the genes can be used to produce chimeric genes for transformation of plants to provide controlled expression of heterologous DNA during fruit development, or in response to exogenous developmental signals, such as ethylene signals. The heterologous protein, e.g. a therapeutic protein, can be isolated from the fruit or consumed directly in the transformed fruit. The identity of

## ALIGNMENTS

CC the present sequence is not given. The entire chitinase DNA sequence has  
CC been decoded in all 3 reading frames to provide the amino acid sequences  
CC given in AAY05845-47  
XX  
SQ Sequence 4924 BP; 1304 A; 1154 C; 1037 G; 1425 T; 0 U; 4 Other;

Query Match 98.4%; Score 2120.8; DB 2; Length 4924;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2149; Conservative 0; Mismatches 3; Indels 5; Gaps 2;

QY 1 GGAATCCAACTTTAGGAATGATCTTAAATATTTAGTATTAAGTTCAGAGTTAGAAAAA 60  
DB |||||  
QY 1 GGAATCCAACTTTAGGAATGATCTTAAATATTTAGTATTAAGTTCAGAGTTAGAAAAA 60  
DB |||||

QY 61 TCTTTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTGAAACGGTGTA CATGTCTCCGA 120  
DB |||||

QY 121 TGGACTCAGCTTGGTTTCAATTCGGNAAGTTTCGAAGAGTGCATAAGAAATATTTGATTGG 180  
DB |||||

QY 121 TGGACTCAGCTTGGTTTCAATTCGGNAAGTTTCGAAGAGTGCATAAGAAATATTTGATTGG 180  
DB |||||

QY 181 ATTCTTTCACTCGGTGGTCTTCATGAGTGAACCTCAAGAGTCCCTCCAAATATCAAAAG 240  
DB |||||

QY 181 ATTCTTTCACTCGGTGGTCTTCATGAGTGAACCTCAAGAGTCCCTCCAAATATCAAAAG 240  
DB |||||

QY 241 CCGAATCACAAATGGAATGATGATTCATTTTGTCTTAATGCAAAACAGGGCAT 300  
DB |||||

QY 241 CCGAATCACAAATGGAATGATGATTCATTTTGTCTTAATGCAAAACAGGGCAT 300  
DB |||||

QY 301 TCATAGCCTTTGTCTTTAAAGCAAAACATTTCTCCGATTCATCCATTCGGTCAATCG 360  
DB |||||

QY 301 TCATAGCCTTTGTCTTTAAAGCAAAACATTTCTCTCCGATTCATCCATTCGGTCAATCG 360  
DB |||||

QY 361 GAAAGAGAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGAAATCCATGAAA 420  
DB |||||

QY 361 GAAAGAGAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGAAATCCATGAAA 420  
DB |||||

QY 421 TGAGGAAGATCCTCATATGAGTTTCCAAATCATGTAAATTCGACTCATTAACATAGTG 480  
DB |||||

QY 481 GATGTGTAATGAATGACCTTCATGCTATCTCTTGGGTATTAACCAAAATATGAGA 540  
DB |||||

QY 481 GATGTGTAATGAATGACCTTCATGCTATCTCTTGGGTATTAACCAAAATATGAGA 540  
DB |||||

QY 541 GTGAGCCTTGTCTCTGATACCAATTTGTTAGGATCAGATGGCACTAAGAGAGGGGGAGT 600  
DB |||||

QY 541 GTGAGCCTTGTCTCTGATACCAATTTGTTAGGATCAGATGGCACTAAGAGAGGGGGAGT 600  
DB |||||

QY 601 GAAATTAGTGCAGTGAATTAACCTTATTAAGTTTAAATGAATTCGTAATACGAGAAGA 660  
DB |||||

QY 601 GAAATTAGTGCAGTGAATTAACCTTATTAAGTTTAAATGAATTCGTAATACGAGAAGA 660  
DB |||||

QY 661 TTTTCGTTTAAATAGTAATCTTGAGTAGAGTAAACCAAAAGTTTAAAGTAGTAAATAC 720  
DB |||||

QY 661 TTTTCGTTTAAATAGTAATCTTGAGTAGAGTAAACCAAAAGTTTAAAGTAGTAAATAC 720  
DB |||||

QY 721 AATTTTCGGGAAGTAAAGTAACTCACATTTCAAGGAACATACCAATTTAAAGTTTCGGT 780  
DB |||||

QY 721 AATTTTCGGGAAGTAAAGTAACTCACATTTCAAGGAACATACCAATTTAAAGTTTCGGT 780  
DB |||||

QY 781 CAAAATGACCTACATCCACTTGTGAAGCCTTTCGAAGAGGCTCCGAACCTCCACTAGC 840  
DB |||||

QY 781 CAAAATGACCTACATCCACTTGTGAAGCCTTTCGAAGAGGCTCCGAACCTCCACTAGC 840  
DB |||||

QY 841 AAATCACTTTGAAGGGAGGAGCAATACCTCTCTTAACNACCTTTTACAATGGTTTCATAC 900  
DB |||||

QY 841 AAATCACTTTGAAGGGAGGAGCAATACCTCTCTTAACNACCTTTTACAATGGTTTCATAC 900  
DB |||||

QY 901 TCTTACAAATTTTCAACGAGAAGAGGAGGTGAACATGCAAGCAATTTGAAACCAAGACT 960  
DB |||||

QY 901 TCTTACAAATTTTCAACGAGAAGAGGAGGTGAACATGCAAGCAATTTGAAACCAAGACT 960  
DB |||||

QY 961 TGCTAAGACTTTGCTAAGGCTTTTCTCAATCTATTGCTTCTCTCAAAAGTTGTATCT 1020  
DB |||||

QY 961 TGCTAAGACTTTGCTAAGGCTTTTCTCAATCTATTGCTTCTCTCAAAAGTTGTATCT 1020  
DB |||||

QY 1021 CTGCTCAGAAATTTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080  
DB |||||

QY 1021 CTGCTCAGAAATTTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080  
DB |||||

QY 1081 TCGAATGCTCTTGGGTTCCGAGGTTGCCGGTCCGACCGCTCTCAGTGTTCGACACTGG 1140  
DB |||||

QY 1081 TCGAATGCTCTTGGGTTCCGAGGTTGCCGGTCCGACCGCTCTCAGTGTTCGACACTGG 1140  
DB |||||

QY 1141 ACAGTGTACTAGCGGTGCGACCCCGGACCTCTCGGGTGTGGCGGTGCGACCGCCTAG 1200  
DB |||||

QY 1141 ACAGTGTACTAGCGGTGCGACCCCGGACCTCTCGGGTGTGGCGGTGCGACCGCCTAG 1200  
DB |||||

QY 1201 ACTTTTTCACTCAGTGTGGATTTCCAACTTTGAGACCCAAACAGTCCGAGTCC 1260  
DB |||||

QY 1201 ACTTTTTCACTCAGTGTGGATTTCCAACTTTGAGACCCAAACAGTCCGAGTCC 1260  
DB |||||

QY 1261 AATTGACCCGTAACCGGATTTATAGGATTAACCTTAATCCTTAACCTTAATATATGAAA 1320  
DB |||||

QY 1261 AATTGACCCGTAACCGGATTTATAGGATTAACCTTAATCCTTAACCTTAATATATGAAA 1320  
DB |||||

QY 1321 CTAGCAACTGAAAAATATAGTCTTAAGCAAGTTTTTAAACCGCAAAAGCTCGAGTCTTCT 1380  
DB |||||

QY 1321 CTAGCAACTGAAAAATATAGTCTTAAGCAAGTTTTTAAACCGCAAAAGCTCGAGTCTTCT 1380  
DB |||||

QY 1381 CCGCGCATCTTTGGCGAGACTTCGATATACCTTTGGGATTTCTTCTAGCGGACTCCTAGT 1440  
DB |||||

QY 1381 CCGCGCATCTTTGGCGAGACTTCGATATACCTTTGGGATTTCTTCTAGCGGACTCCTAGT 1440  
DB |||||

QY 1441 AGGTCGCCATCTTCTGCGGAGTTTACCGGAGTTCAGCAAGCTTCTCGGTGATCTCCGCA 1500  
DB |||||

QY 1441 AGGTCGCCATCTTCTGCGGAGTTTACCGGAGTTCAGCAAGCTTCTCGGTGATCTCCGCA 1500  
DB |||||

QY 1501 ACCGCCGATGATCTTCTGCGGAGACTTTTCGAAAACTTCGACCAAGTCCCGGATTTCTCTC 1560  
DB |||||

QY 1501 ACCGCCGATGATCTTCTGCGGAGACTTTTCGAAAACTTCGACCAAGTCCCGGATTTCTCTC 1560  
DB |||||

QY 1561 GGTTCGTCGACAGCATCTCTAACGAAACTTCGGAATCTCTTGAATGTCCATCGAACTTG 1620  
DB |||||

QY 1561 GGTTCGTCGACAGCATCTCTAACGAAACTTCGGAATCTCTTGAATGTCCATCGAACTTG 1620  
DB |||||

QY 1621 ACTCCGTAGGCTTGTCTTATTTTTCAGGCTATCATAGTTAATCTTACATCTTAATCTC 1680  
DB |||||

QY 1621 ACTCCGTAGGCTTGTCTTATTTTTCAGGCTATCATAGTTAATCTTACATCTTAATCTC 1680  
DB |||||

QY 1681 AATAATATGATTAAGTTAATTAACCCATCAATTTGAATTTTCATCATCAAAATTCGACATTC 1740  
DB |||||

QY 1681 AATAATATGATTAAGTTAATTAACCCATCAATTTGAATTTTCATCATCAAAATTCGACATTC 1740  
DB |||||

QY 1741 AACAAACATCCGTACTCAATAACCCATCAGGCTATAGTTACGTGACTATCTACTGTGATC 1800  
DB |||||

QY 1741 AACAAACATCCGTACTCAATAACCCATCAGGCTATAGTTACGTGACTATCTACTGTGATC 1800  
DB |||||

QY 1801 CGTACGTGAAGTTAGCGGATCATGATCCAGTTCGTTGACATTTATGGCCGAACAGTATC 1860  
DB |||||

QY 1801 CGTACGTGAAGTTAGCGGATCATGATCCAGTTCGTTGACATTTATGGCCGAACAGTATC 1860  
DB |||||

QY 1861 CTTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCTACCGCTCTCTTTTTTATTAATCT 1920  
DB |||||

QY 1861 CTTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCTACCGCTCTCTTTTTTATTAATCT 1920  
DB |||||

QY 1921 TTGAAAGAAATTAATCAAAACAGATACAAATTAACAGGTGAGACATCTGACATGCTA 1980  
DB |||||

QY 1921 TTGAAAGAAATTAATCAAAACAGATACAAATTAACAGGTGAGACATCTGACATGCTA 1980  
DB |||||

QY 1981 GTCTCTGAAAGCAATTAATTCGCGCATCCACAGAGCTGCTCAGCTTTCATACCCACTTTT 2040  
DB |||||

QY 1981 GTCTCTGAAAGCAATTAATTCGCGCATCCACAGAGCTGCTCAGCTTTCATACCCACTTTT 2040  
DB |||||

QY 2041 TCCTACAT-ACCATTGCGCATGGCTTTGTTGATGACAGACACCAAGCTTGCTTTGG 2099  
 |||||  
 Db 2041 TCCTACATAACCATGTGCGCATGGCTTTGTTGATGACAGACACCAAGCTTGCTTTGG 2100  
 |||||  
 QY 2100 TTGTGCTTAACAGAGAGAGAGAGAGAGACGATAGCTTCTCATTCACATATGG 2156  
 |||||  
 Db 2101 TTGTGCTTAAC-AGAGAGAGAGAGAGACGATAGCTTCTCATTCACATATGG 2153  
 |||||

RESULT 2  
 AAV28690  
 ID AAV28690 standard; cDNA; 809 BP.  
 AC AAV28690;  
 XX  
 DT 29-JUL-1998 (first entry)  
 XX  
 DE Ripening banana pulp cDNA clone U-D86 SEQ ID NO:48.  
 XX  
 KW Banana; ripening; pulp; Musa acuminata cv. Grand Nain; fruit;  
 KW genetic control; tissue senescence; ss.  
 XX  
 OS Musa acuminata.  
 XX  
 PN WO9811228-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PF 08-SEP-1997; 97WO-GB002424.  
 XX  
 PR 10-SEP-1996; 96GB-00018862.  
 PR 25-APR-1997; 97GB-00008366.  
 XX  
 PA (ZENE ) ZENECA LTD.  
 XX  
 PI Seymour GB, Bird CR, Medina-Suarez RDJ;  
 XX  
 DR WPI; 1998-207389/18.  
 XX  
 PT Modulation of ripening or tissue senescence in bananas - comprises use of  
 PT DNA isolated from ripening banana pulp to produce genetically modified  
 PT fruit.  
 XX  
 PS Claim 1; Page 48-49; 72pp; English.  
 XX

CC The present sequence represents a cDNA clone isolated from ripening  
 CC banana pulp. 57 clones were isolated and are given in AAV28643 to  
 CC AAV28699. The cDNA clone sequences can be used in a method of modulating  
 CC ripening or tissue senescence process in plants of the genus Musa. The  
 CC method comprises: (a) inserting into the plant material at least 1 of the  
 CC 57 sequences (as above); (b) regenerating the plant material; and (c)  
 CC selecting from the transformed regenerants, plants with modulated  
 CC ripening or tissue senescence characteristics. Also described in the  
 CC present invention are: (1) plants, their progeny, seed and material  
 CC obtained from the plants, produced by the above method; (2) a vector  
 CC functional in plants comprising a promoter region which is operably in  
 CC plant cells, a polynucleotide sequence as defined above, and a  
 CC transcription termination sequence; and (3) a method of controlling plant  
 CC pathogens comprising the application of anti-pathogenic agent to plants  
 CC of (1)  
 XX

SQ Sequence 809 BP; 191 A; 233 C; 197 G; 180 T; 0 U; 8 Other;  
 Query Match 3.5%; Score 75; DB 2; Length 809;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2082 CCACAGCTTGCTTTGGTTGTCCTAACAGAGAGAGAGAGAGACACCGATAGCCT 2141  
 |||||  
 Db 9 CCACAGCTTGCTTTGGTTGTCCTAACAGAGAGAGAGAGAGACACCGATAGCCT 68  
 |||||  
 QY 2142 CTTCAATTCATATGG 2156  
 |||||

Db 69 CCTCATTCACATATGG 83  
 RESULT 3  
 AAV22222/c  
 ID AAV22222 standard; cDNA; 2417 BP.  
 XX  
 AC AAV22222;  
 XX  
 DT 23-JUL-1998 (first entry)  
 XX  
 DE SIRE-1 cDNA from lambda gtl1 soybean cDNA library.  
 XX  
 KW SIRE-1; Gm776 fragment; soybean; retrotransposon; copia; Ty1; ss.  
 XX  
 OS Glycine max.  
 XX  
 PN WO9809505-A1.  
 XX  
 PD 12-MAR-1998.  
 XX  
 PF 25-AUG-1997; 97WO-US014802.  
 XX  
 PR 09-SEP-1996; 96US-0025853P.  
 XX  
 PA (LOYO ) UNIV LOYOLA CHICAGO.  
 XX  
 PI Laten HM;  
 XX  
 DR WPI; 1998-193245/17.  
 XX  
 PT New plant retrovirus sequence, SIRE-1 - useful for, e.g. genetic  
 PT engineering in plants.  
 XX  
 PS Claim 1; Fig 6; 199pp; English.  
 XX

CC This sequence represents the SIRE-1 cDNA sequence, isolated from a  
 CC lambda gtl1 soybean cDNA library, and is a polynucleotide of the  
 CC invention. SIRE-1 is a member of the copia/Ty1 retrotransposon family.  
 CC SIRE-1 can be used for genetic engineering in plants, e.g. for conferring  
 CC resistance to disease, insect infestation, herbicide tolerance, enhance  
 CC N2 fixation, or enhanced vigour  
 XX  
 SQ Sequence 2417 BP; 807 A; 460 C; 596 G; 554 T; 0 U; 0 Other;  
 Query Match 2.9%; Score 62.8; DB 2; Length 2417;  
 Best Local Similarity 50.6%; Pred. No. 7.6e-07;  
 Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;  
 QY 4 TCCCAACCTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAGTTAGAAAATCT 63  
 |||||  
 Db 786 TCAAATCTCTTAGGCAAGGATCTGAGATCTTTCTCACCAGCTTTTTCATCAGTCATCTT 727  
 |||||  
 QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATC-CGTGAAACGGTGACATGCTCCGATG 122  
 |||||  
 Db 726 TCTCCCAAGGCACTGCAAGCATTTGCAATTTCAAGATGTTGATGAAAGTCATGAATA 667  
 |||||  
 QY 123 GACTCACTTGGTTTCATTTCGGAAAAAGTTGGAAGAGTGCATAGAAATATTTGGAT 182  
 |||||  
 Db 666 CACTCTCTCTCTTCATCTTCAGATTTTCGAATTTTGTGGCCAAATAGTTGCAATCTGGAC 607  
 |||||  
 QY 183 TCTTTCACTCGTGTGGTTCCTTCATGAGTGCCTCAAGAGTCTTCCAAATATCAAAAGCC 242  
 |||||  
 Db 606 ATCTTCACTTTGGAGGTTCTTTCATGAGTGGTTTTCAGGATCTCCCATGATCTCTTGGCC 547  
 |||||  
 QY 243 GAATCACAATTTGAATGTGATTTGATTTTCTATTTTGTCTAAATGCAAAACAGGCAATC 302  
 |||||  
 Db 546 ACTGTGATGTGTTGATCAGTCTGAAGATATTTCTGTCAACTCCATTTGAATAGACATTC 487  
 |||||  
 QY 303 ATAGCCTTTGTGTTTAAAGCAAAAACATTTCTCTCCGATTTCAATCCATTC 352  
 |||||  
 Db 486 AAAGCTTTGGAGTTTCCAGTCCCAATTCGTCCTTCTTTAGTCCAGTC 437  
 |||||

RESULT 4  
 AAF22305\_02  
 Continuation (3 of 11) of AAF22305 from base 200001 (Arabidopsis thaliana chromosome 4)  
 WP Sequence split into 11 fragments LOCUS AAF22305 Accession Aaf22305  

WP	Fragment Name	Begin	End
WP	AAF22305_00	1	110000
WP	AAF22305_01	100001	210000
WP	AAF22305_02	200001	310000
WP	AAF22305_03	300001	410000
WP	AAF22305_04	400001	510000
WP	AAF22305_05	500001	610000
WP	AAF22305_06	600001	710000
WP	AAF22305_07	700001	810000
WP	AAF22305_08	800001	910000
WP	AAF22305_09	900001	1010000
WP	AAF22305_10	1000001	1082138

Query Match 2.5%; Score 54; DB 3; Length 110000;  
 Best Local Similarity 51.7%; Pred. No. 0.0018;  
 Matches 123; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY	114	TCTCCGATGGACTCACCTGGTTTCATTCGGAAAAGTTCGAAAGAGTGCATAGAAATATTG	173
DB	31784	TCCTCAATGTTCTCTGTTTCTTCATCTAGATTTCAACTGAGATGCTACATCTCG	31843
QY	174	ATTTTGATTCCTTTCACCTCGGTTGGTCCCTCATGAGTGACCTCAAGAGTCCTCCAAATA	233
DB	31844	ATTCGGATCTTTTGACACTACTTGCTCCCTTCATATGCTTTAGCAAGTTTATCCCATGCT	31903
QY	234	TCAAAAGCCGAATCACAAATTAATGATGATGAATTCATTTTGTCTAATGCACAAAAC	293
DB	31904	TCCTTAGCTGATTCACACTGTTGAATCCGCTTGAAITGATTTGATTCAGAGTTGGAAT	31963
QY	294	AGGCGATTCATAGCCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCAT	351
DB	31964	ATCAAGACAGTGCCTTGAATTAGCTTTGGCCTTTGCTCTTCGCAATCATCCAT	32021

RESULT 5  
 AAF22305\_03  
 Continuation (4 of 11) of AAF22305 from base 300001 (Arabidopsis thaliana chromosome 4)  
 WP Sequence split into 11 fragments LOCUS AAF22305 Accession Aaf22305  

WP	Fragment Name	Begin	End
WP	AAF22305_00	1	110000
WP	AAF22305_01	100001	210000
WP	AAF22305_02	200001	310000
WP	AAF22305_03	300001	410000
WP	AAF22305_04	400001	510000
WP	AAF22305_05	500001	610000
WP	AAF22305_06	600001	710000
WP	AAF22305_07	700001	810000
WP	AAF22305_08	800001	910000
WP	AAF22305_09	900001	1010000
WP	AAF22305_10	1000001	1082138

Query Match 2.5%; Score 53.4; DB 3; Length 110000;  
 Best Local Similarity 50.2%; Pred. No. 0.0026;  
 Matches 132; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY	107	GTACATGTCCTCGATGGACTCACCTGGTTTCATTCGGAAAAGTTCGAAAGAGTGCATAAG	166
DB	14372	GTTGAAGTCATCCCGGATCTTCAGCTTCATGGTCAGATTCCTCAACTCTGACGCAAG	14431
QY	167	AATATTGATTTTGGATTCCTTCACTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCT	226
DB	14432	CATGTCCAGACGAGTACGCTTCACATTTGTTGTCATTCAAATGAGACTTCGAGAAATTC	14491
QY	227	CCAAATATCAAAAGCCGAATCACAAATTAATGATGATGAATTCATTTTGTCTAATGC	286
DB	14492	CCATACTTCTCTTGTCTGATAGCATCTTGAACCTGAGTGAACTTATTCCTTAACAAGA	14551
QY	287	ACAAACAGGCGATTCATAGCCTTTGTGTTTAAAGCAAAACATTTCTCTCCGATTCATC	346









XX AC F73552;  
XX DT 20-NOV-2003 (first entry)  
XX DE Staphylococcus aureus DNA #1232.  
XX KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
XX KW enzymatic assay; antibiotic target; gene; ds.  
XX OS Staphylococcus aureus.  
XX PN WO200294868-A2.  
XX PD 28-NOV-2002.  
XX PF 27-MAR-2002; 2002WO-IB002637.  
XX PR 27-MAR-2001; 2001GB-00007661.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Masignani V, Mora M, Scarselli M;  
XX WPI; 2003-120786/11.  
XX P-PSDB; ABM71992.  
XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
XX preventing Staphylococcal infection, specifically an infection caused by  
XX S. aureus, e.g. sepsis.  
XX Claim 6; SEQ ID NO 2463; 49pp; English.  
XX The invention relates to novel genes and encoded proteins from  
XX Staphylococcus aureus. A composition comprising the S. aureus protein, a  
XX nucleic acid encoding the protein, or an antibody to the protein, is  
XX useful as a pharmaceutical, particularly as a vaccine for treating or  
XX preventing infection due to Staphylococcus bacteria, specifically an  
XX infection caused by S. aureus. The composition is particularly useful for  
XX treating or preventing sepsis in a patient. The composition can also be  
XX used for diagnostics. The protein is also used in an assay for enzymatic  
XX studies and as a target for antibiotics. This sequence represents one of  
XX the novel S. aureus genes of the invention  
XX SQ Sequence 1722 BP; 657 A; 240 C; 310 G; 515 T; 0 U; 0 Other;  
Query Match 1.9%; Score 41.2; DB 7; Length 1722;  
Best Local Similarity 49.1%; Pred. No. 0.96;  
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 564 TGTAGGATCAGAGTGGCACTAAGAGAGAGGGGGAGTGAATTAGTGCAGTGAATAAAC 623  
DB 1293 TGTTCCTTAGAAGTCCCAATAGGTACGGAAGTATTTGAAGTGGCTGTGGTAATCAACG 1352  
QY 624 TTATAGCTTAAAGTAAATCGTAAATACGAGAGATTCGTTTTATAGTAACCTCGAG 683  
DB 1353 TAATAAATATGATATATCTGCAGAAACAGTGCATTAATTCATTTGAATATGATTTTA 1412  
QY 684 TAGATGAACAAACAAAGTTAAACAGTAGTGTAATAACAAATTCGGGAAAGTAAGACTCA 743  
DB 1413 TAGATTACACCGTATTTCCACAAAGACTTTTAATACATTCGTTATATCTTACAGCTAT 1472  
QY 744 CACATTCAGGAACATACCAATTTAAAGTGGTTCGGTCAAAA 785  
DB 1473 TACATTAACGATTCATCAATCTCATTTGAAGTTAAAGGTAAAAA 1514  
RESULT 13  
AAV74565  
ID AAV74565 standard; DNA; 2881 BP.  
XX AAV74565;  
XX AC AAV74565;  
XX

DT 16-MAR-1999 (first entry)  
XX DE Staphylococcus aureus contig SEQ ID #254.  
XX KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
XX KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
XX KW skin infection; surgical wound infection; scalded skin syndrome;  
XX KW toxic shock syndrome; ds.  
XX OS Staphylococcus aureus.  
XX FH Key Location/Qualifiers  
FT misc\_feature 541..600  
FT /\*tag= a  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They are  
FT included to maintain the nucleotide numbering given in  
FT the specification for this DNA sequence"  
FT 2341..2400  
FT misc\_feature  
FT /\*tag= b  
FT /note= "these bases represent a line of missing text in  
FT the sequence listing in the specification. They are  
FT included to maintain the nucleotide numbering given in  
FT the specification for this DNA sequence"  
XX EP786519-A2.  
XX 30-JUL-1997.  
XX PF 07-JAN-1997; 97EP-00100117.  
XX PR 05-JAN-1996; 96US-0009861P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Kunsch CA, Choi GH, Barash SC, Dillon PJ, Fannon MR, Rosen CA;  
XX WPI; 1997-374922/35.  
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -  
XX stored on computer readable medium and used in the production of anti-  
XX S.aureus vaccines.  
XX Claim 1; Page 1047-1049; 3271pp; English.  
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
XX of the invention. The DNA sequences are recorded on a computer readable  
XX medium, preferably selected from a floppy or hard disk, random access  
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
XX the S.aureus DNA sequences allows putative functions to be assigned so  
XX that protein-encoding or regulatory regions of commercial, therapeutic or  
XX industrial importance can be obtained. Specifically, sequences which are  
XX likely to encode antigens have been identified and these polypeptides can  
XX be used in a vaccine composition against S.aureus infection. The  
XX polypeptides can also be used in a kit for the immunodetection of  
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
XX skin and surgical wound infections, scalded skin syndrome, toxic shock  
XX syndrome, etc. Organisms transformed with the DNA sequences can be used  
XX for recombinant production of the polypeptides. The new DNA sequences  
XX (and their fragments) are useful as primers or probes for isolating  
XX homologues of any of the S.aureus DNA sequences contained on the computer  
XX readable medium  
XX Sequence 2881 BP; 1049 A; 376 C; 505 G; 826 T; 0 U; 125 Other;  
SQ Query Match 1.9%; Score 41.2; DB 2; Length 2881;  
Best Local Similarity 49.1%; Pred. No. 1.2;  
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;  
QY 564 TGTAGGATCAGAGTGGCACTAAGAGAGGGGGAGTGAATTAGTGCAGTGAATAAAC 623  
DB 1481 TGTTCCTTAGAAGTCCCAATAGGTACGGAAGTATTTGAAGTGGCTGTGGTAATCAACG 1540



Mon Sep 20 12:58:17 2004

Db 5082 CAACAACTACTACTAAATTCGAAATTAATAAATTACTAAATTTAAATC 5033

Search completed: September 18, 2004, 19:20:43  
JOB time : 765.234 secs



```

RESULT 2
US-08-282-581-5
; Sequence 5, Application US/08282581
; Patent No. 5670349
; GENERAL INFORMATION:
; APPLICANT: Cramer, Carole L.
; APPLICANT: Weissenborn, Deborah L.
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND
; TITLE OF INVENTION: POST-HARVEST PRODUCTION OF GENE PRODUCTS
; TITLE OF INVENTION: IN PLANTS AND PLANT CELL CULTURES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,581
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18872
; REFERENCE/DOCKET NUMBER: 7956-009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-282-581-5

Query Match 2.4%; Score 52.6; DB 1; Length 515;
Best Local Similarity 51.5%; Pred. No. 5.2e-06;
Matches 121; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 135 TTCATTTCGGAAAAAGTTCCAAAGAGTGCATAAGAATATTGATTTTGGAAATCTTTCACCTCGG 194
DB 122 TTCATTTTGAAGTTCTCATATCGTGAGTGAGCATGTCAATCTCTGGATTTCTTTGACTTGT 181
QY 195 TTGGTGCCTTCATGAGTGACCTCAAGAGTCTTCAATATCAAAAGCCGAATCACAAATT 254
DB 182 TCAGTTTCTTCATGTGTAGTCAACAAGCAATCCCAAGATTTCTTTAGCAGACTCACAGCT 241
QY 255 GAAATGTGATGTAATTCATTTTCTTAATGCAAAAACAGGGGCATTCATAGCCTTTTGTG 314
DB 242 GACACTCTATTGTACTCATCAGGTCCTATCCCACAGACCATAAGAGTTTACCTTTGAAG 301
QY 315 TTTAAGCAAAAACATCTCTTCGCAATTCATCCCATTCGCTCATCGGAAGAGAAA 369
DB 302 CCCTTTTCTATCTTTTTCCTGTGCAGCATCATCAATTTCTGCTGGGCTTTGGAA 356

RESULT 3
US-08-550-544-5
; Sequence 5, Application US/08550544
; Patent No. 5689056
; GENERAL INFORMATION:
; APPLICANT: Cramer, Carole L.
; APPLICANT: Weissenborn, Deborah L.
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM

```



```

; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REFERENCE/DOCKET NUMBER: 29,768
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptz9pt-Fls
; IMMEDIATE SOURCE:
; US-08-232-463-14

Query Match 2.1%; Score 48.4; DB 1; Length 7218;
Best Local Similarity 12.0%; Pred. No. 0.00052;
Matches 61; Conservative 214; Mismatches 235; Indels 0; Gaps 0;

QY 1454 TGTGGGAGTTTACGAGTAGCGAACCTTCTCGGTGATCTTCGCAACCGCGGATGATC 1513
Db TGGGCTACTATACTATTTCTCTGTTGGCATACGCTCACAGAAATTAATTCGAGC 1038

QY 1514 TCTTCGCGACACTTTCGAAACTTCGAAAGTCCCGGATTTCTTCGGTTGGTTCGGAC 1573
Db TGGGCTACTATACTATTTCTCTGTTGGCATACGCTCACAGAAATTAATTCGAGC 1038

QY 1039 TTGGCTGAGGTTCGAGGAGCTTCGGATTTTTCTCTGTTGGCATACGCTCACAGAAATTAATTCGAGC 1038
Db TGGGCTACTATACTATTTCTCTGTTGGCATACGCTCACAGAAATTAATTCGAGC 1038

QY 1574 AGCATCTCAAGAACTTCGAGTCTCGTCTGATCTTCGCAACCGCGGATGATC 1513
Db TGGGCTACTATACTATTTCTCTGTTGGCATACGCTCACAGAAATTAATTCGAGC 1038

QY 1099 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1158
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1158

QY 1634 TGCTTTATATTTTCAGGCTATCATAGTTAATCTCATACATCTTAATCAATATGATTT 1593
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1218

QY 1694 AGAATAATTAACCATCAATGATTTATCATCAAAATTCGACATCAACAAATCCGCT 1753
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1278

QY 1754 ACTCAATACCATCAGGCTATAGTACGAGTATCTATCTGATCGGACGATGAGTT 1813
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1338

QY 1814 AGCGAGTCATGATCCAGGCTGCTCACTATTTCGCGGACACATGATCCCTATCCAAATC 1873
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1398

QY 1874 CAGTCTTCTCACTCTTACGCTACCGCTCTCTTTTATTACTTTTGAAGAAATTCGA 1933
Db YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1458

QY 1934 AATCAAAACAGATACAAAATAACACGGTGA 1963
Db ACTACTTGATAGATAGTAAATACAGTGA 1488
```

```

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOMLOPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Hardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REFERENCE/DOCKET NUMBER: 29,768
; REGISTRATION NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptz9pt-Fls
; US-08-232-463-14
```

```

Query Match 2.1%; Score 45; DB 1; Length 7218;
Best Local Similarity 7.0%; Pred. No. 0.0058;
Matches 30; Conservative 212; Mismatches 187; Indels 0; Gaps 0;

QY 320 AGCAAAACATCTTCTCGATTCATCCCATTCGCTCATCGGAGAGAAAATTTTGA 379
Db 1493 AGGCATCACTGTAATTAATTCATCTATGCAAGTAGTTAAAGAGATAGAGAAATTTGCTACR 1434

QY 380 TCCATTTTCGACATAGACCAAGCTCGAAATCCATGGAATCCAGGAATGAGGAAGATCCTCATATG 439
Db 1433 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1374

QY 440 AGTTTTCATATCATGTAATTCGACTCATTAACATAGGTGGATGTTAATGAAATGACC 499
Db 1373 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1314

QY 500 CTCATGCSCTATCTCTTGGGTATTAACCAAAATATGAGAGTACGCTTGTCTCTGATAC 559
Db 1313 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1254

QY 560 CAATTCTTAGGATCAGATGCGCACTAAGAGAGGGGGAGTGAATAGTGCAGTGAATTA 619
Db 560 CAATTCTTAGGATCAGATGCGCACTAAGAGAGGGGGAGTGAATAGTGCAGTGAATTA 619
```



GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 17  
; LENGTH: 5535  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-17

Query Match 1.7%; Score 37.6; DB 4; Length 5535;  
Best Local Similarity 57.8%; Pred. No. 0.92;  
Matches 67; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1615 AACTTGACTCGGTAGCTTGTATATTTTCAGGCTATCAGTTAAATCCTCATACT 1674  
Db 2300 AACTTACTTACTCAATACCTTTTCTTTCCCTCAATTAATAATCCAAATTC 2241  
QY 1675 TAACTCAATATATGATTAATTAATCAATCAATCAATCAATCAATCAATCA 1730  
Db 2240 AAATAAATACGATACTCATCTATATTAATCCCAATTAATAAATCCAAATTA 2185

RESULT 9  
US-10-204-708-11  
; Sequence 11, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 11  
; LENGTH: 6317  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-11

Query Match 1.7%; Score 37; DB 4; Length 6317;  
Best Local Similarity 48.8%; Pred. No. 1.5;  
Matches 100; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 564 TGTTAGGATCAGAGTGGCCTAAGAGAGGGGGGAGTGAATTAGTCGAGTCGATTAAAC 623  
Db 2087 TGGTGGTTTATGTTTGTAAATTAGTAATTCGGGAGGTAGAGTTAGAGGATTATTGAGT 2146  
QY 624 TTATAAGTTTAAATAAGTTCGTAAATACGAGAAGATTTGTTTTTAATAGTAATTTGAG 683  
Db 2147 TTAAGAGTTTGGGATTTAGTTTGGTAAATATAGGAAGATTTATTTTATAAAAAATTA 2206  
QY 684 TAGATGAACCAACCAAGTTAACAGTAGCTGTAATAAACAATTCGGGAAGTAAGAACTCA 743  
Db 2207 AAAAAAATAATTAGTGGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2266  
QY 744 CACATTCAGGAACATACCAATTTA 768  
Db 2267 GAGTAGGAGAATTTTATTTAGTTTA 2291

RESULT 10  
US-09-621-976-8976  
; Sequence 8976, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 8976  
; LENGTH: 399  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-8976

Query Match 1.7%; Score 36.6; DB 4; Length 399;  
Best Local Similarity 10.0%; Pred. No. 0.37;  
Matches 30; Conservative 141; Mismatches 130; Indels 0; Gaps 0;  
QY 575 GAGTGCCTAAGAGAGGGGGGAGTGAATTAGTCGAGTCGATTAAATAAGTTTA 634  
Db 44 RRRRRRAWWKSMCMWKKSKSWRSWGMWTKEMKGRGAASWAGYMSWMTYTRWRYR 103  
QY 635 AAAATGAATTCGTAATACAGAGAAGATTTGTTTTTAATAGTAATCTGAGTAGATAAAC 694  
Db 104 YRKCACTKWRAAGMWGKAGWAWAYAKWYMAWRRTAMKYWAMKSKSRMRRAWYAWM 163  
QY 695 CAAAAGTTAACAGTAGTGTAAATAACAATTCGGGAAGTAAGAACTCACACATTCAGG 754  
Db 164 YWMAARTWGRASCYRGAYMASAGYMYMMYMMYMMYMMYMMYMMYMMYMMYMMYMMY 223  
QY 755 AACATACCAATTTAAAGTGGTTCGGTCAAAAATGACCTACATCCACTTGTGAAGCCCTCT 814  
Db 224 CWSSYCMWGAQWYMYWKTSEWSYVYSSYRKYRSCCWSMSYWKTYRWSYSCSY 283  
QY 815 CGAAGAGGCTCCCACTCCACTAGCAATACATCTTTGAAGGGGAGGACAAATACCTCTC 874  
Db 284 SYKTRASCMMCCMKWKKWMMYMMYMMYMMYMMYMMYMMYMMYMMYMMYMMYMMYMM 343  
QY 875 T 875  
Db 344 W 344

RESULT 11  
US-09-023-655-768/c  
; Sequence 768, Application US/09023655

```

; Patent No. 660789
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 768:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTT01
; CLONE: 602872
; US-09-023-655-768

Query Match 1.7%; Score 36.6; DB 4; Length 2871;
Best Local Similarity 45.5%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 151;

QY 510 ATCTCTCTGGGTATTAAACCAATATGAGGTGAGCCTTGCTCTGATACCAATTTGTAG 569
Db 717 ATATTACAGGTCAAGAAACAAAGCTGCTGCCAGTCATGTTTGGCAATAACGTTTG 658
QY 570 GATCAGAGTGGCACTAAGAGAGGGGGAGTGAATAGTGCAGTGGATTAACCTATAA 629
Db 657 GGGTCAGCGGAAAAAGGGGAGAAAGGAAAGGAGGAGGAGAAATAACTAACPTT 598
QY 630 GTTTAAAAATGAATTCCTTAATACGAGAAGATTTCGTTTAAATAGTAACCTGAGTAGATG 689
Db 597 CTGAAACACAACTTGGCTTAACTCGCAAAATAAGGCTTTTCGGGAGAAATGAAAGCCT 538
QY 690 AAACCAAAAGTTAACAGTGTGTAATTAACAACTTTCCGGGAAAGTAAGACTCACAT 749
Db 537 ATATCAGGATTTAGGTGTGCAATAAACAACACAGCTGACACACAGCAATATCTTACATCAT 478
QY 750 CAAGGAACATACCAATTTAAGTGGTTCGTCACAAAT 786
Db 477 TAATANCTTCTCNATTTATAGATCTATCGTCACAAAT 441

RESULT 12
; Sequence 124-000C-1552
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 768:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTT01
; CLONE: 602872
; US-09-023-655-768

Query Match 1.7%; Score 36.6; DB 4; Length 2871;
Best Local Similarity 45.5%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 151;

QY 510 ATCTCTCTGGGTATTAAACCAATATGAGGTGAGCCTTGCTCTGATACCAATTTGTAG 569
Db 717 ATATTACAGGTCAAGAAACAAAGCTGCTGCCAGTCATGTTTGGCAATAACGTTTG 658
QY 570 GATCAGAGTGGCACTAAGAGAGGGGGAGTGAATAGTGCAGTGGATTAACCTATAA 629
Db 657 GGGTCAGCGGAAAAAGGGGAGAAAGGAAAGGAGGAGGAGAAATAACTAACPTT 598
QY 630 GTTTAAAAATGAATTCCTTAATACGAGAAGATTTCGTTTAAATAGTAACCTGAGTAGATG 689
Db 597 CTGAAACACAACTTGGCTTAACTCGCAAAATAAGGCTTTTCGGGAGAAATGAAAGCCT 538
QY 690 AAACCAAAAGTTAACAGTGTGTAATTAACAACTTTCCGGGAAAGTAAGACTCACAT 749
Db 537 ATATCAGGATTTAGGTGTGCAATAAACAACACAGCTGACACACAGCAATATCTTACATCAT 478
QY 750 CAAGGAACATACCAATTTAAGTGGTTCGTCACAAAT 786
Db 477 TAATANCTTCTCNATTTATAGATCTATCGTCACAAAT 441

RESULT 13
; Sequence 13, Application US/10204708
; Patent No. 667731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 13
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-204-708-13

Query Match 1.7%; Score 36.2; DB 4; Length 6113;

```

```

Best Local Similarity 49.2%; Pred. No. 2.6;
Matches 95; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 591 GGGGGGAGTGAATAGTCGAGTGAATTAAGTTTAAATGAAATCGTAA 650
Db 906 GGTGTGTTTAAATATATATATATATATATATATATATATATAT 965

QY 651 TACGAGAAGATTTCGTTTAAATAGTAACTTCGAGTAGATGAAACCAAAAGTTTACAGTAG 710
Db 966 GTTAATTTTATGTTTAAATATAGTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 1025

QY 711 TGTAATAACAAATTCGGGAAAGTAAAGAACTCACACATTCAGGAAACATACCAATTTAAA 770
Db 1026 ATTAAGAGAAAAAATGAGATATTATATATATATATATATATATATATATATATATATATAT 1085

QY 771 GTGGTTCGGTCAA 783
Db 1086 TTGATTATATATA 1098

```

```

RESULT 14
US-09-539-333D-192
; Sequence 192, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essicou, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET 047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 192
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1501
; OTHER INFORMATION: 99-26201-267 : polymorphic base C or G
; FEATURE:
; NAME/KEY: misc binding
; LOCATION: 1502..1520
; OTHER INFORMATION: 99-26201-267.mis1, complement
; FEATURE:
; NAME/KEY: misc binding
; LOCATION: 1481..1500
; OTHER INFORMATION: 99-26201-267.mis2,
; FEATURE:
; NAME/KEY: primer_bind

```

```

; LOCATION: 1749..1767
; OTHER INFORMATION: upstream amplification primer, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1304..1324
; OTHER INFORMATION: downstream amplification primer
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-26201-267 probe
US-09-539-333D-192

Query Match 1.7%; Score 35.8; DB 4; Length 3001;
Best Local Similarity 49.7%; Pred. No. 2.3;
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 321 GCAAAACATTTCTTCGATTTCATCCATTCGCTCATCGAAGAGAAAATTTTGAAT 380
Db 2324 GCTAAAGTTTCTTCATTCCTCAAAATCTGCAAGACAGACAGAAAGAAAATTTACT 2383

QY 381 CCATTTTCGACATAGACCAAGCTCGAAATCCATGGAAATGAGAAAGATCCTCATATGA 440
Db 2384 CCTATCTCAATAGAGTCACTAGCCCTATATATAATATTATCAAGCCAAATCTTGAATTAT 2443

QY 441 GTTTCCCAATACATGTAAATTCGACTCATTAAACATAGTGGATGTGTAATGAATGACCC 500
Db 2444 ATGTTAGAAACATATGATTAGGCTGAATTAATCACTGGAGAGTGAGGCTGACCC 2503

QY 501 TCA 503
Db 2504 TTA 2506

RESULT 15
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; FILE REFERENCE: JII-002CNC
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
US-08-213-419B-3

Query Match 1.7%; Score 35.8; DB 4; Length 6124;
Best Local Similarity 50.3%; Pred. No. 3.5;
Matches 88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 598 AGTGAATTAGTCAGTGGATTAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAA 657
Db 1797 AATGTACACATGTACAATATATATATATATATATATATATATATATATATATATATATAT 1856

QY 658 AGATTTCGTTTAAATAGTAACTTGAGTAGATGAAACCAAACTTAACAGTAGTGAAT 717

```

Db 1857 ATATTTTGAATATTAAATTAATTCCTTAGAATAACTATATAATTTGATAATCCGAATCT 1916  
Qy 718 AACAAATTCGGGAAAGTAAGAACTCACACATTCAGGAACATACCAATTTTAAAGT 772  
Db 1917 TAAAATGTTACAAAATGAGAAAGTAAAAAACTCCTTATTATATATATATATAAT 1971

Search completed: September 18, 2004, 14:23:48  
Job time : 155.359 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:07:43 ; Search time 967.103 Seconds  
(without alignments)  
11252.468 Million cell updates/sec

Title: US-09-892-635A-44

Perfect score: 2156

Sequence: 1 ggatccaaacttttaggaat.....agcctctcattcattg 2156

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq2:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2154.6	99.9	2156	13	US-09-892-635A-44
2	2139	99.2	2160	13	US-09-892-635A-45
3	2125.6	98.6	4924	13	US-09-892-635A-27
4	127.6	5.9	78333	16	US-10-298-122-3
5	92	4.3	5895	16	US-10-298-122-2
6	81.4	3.8	2760	13	US-10-424-599-108338
7	75	3.5	809	9	US-09-966-881-48
8	73.4	3.4	5016	17	US-10-437-963-10782
9	73.4	3.4	5304	17	US-10-437-963-44132
10	71.8	3.3	3843	17	US-10-437-963-10117
11	71.8	3.3	4326	17	US-10-437-963-9179
12	71.8	3.3	4428	17	US-10-437-963-44137
13	71.8	3.3	4840	17	US-10-437-963-9223
14	71.8	3.3	10369	17	US-10-437-963-10605

c 15	70.2	3.3	4002	17	US-10-437-963-9586	Sequence 9586, Ap
c 16	70.2	3.3	4428	17	US-10-437-963-9621	Sequence 9621, Ap
c 17	69.6	3.2	2160	17	US-10-437-963-101688	Sequence 101688, Ap
c 18	68.6	3.2	4884	17	US-10-437-963-9664	Sequence 9664, Ap
c 19	68	3.2	846	17	US-10-437-963-96678	Sequence 96678, A
c 20	67.4	3.1	4716	17	US-10-437-963-10750	Sequence 10750, A
c 21	67	3.1	4125	17	US-10-437-963-9312	Sequence 9312, Ap
c 22	67	3.1	4620	17	US-10-437-963-10128	Sequence 10128, A
c 23	67	3.1	5370	17	US-10-437-963-9697	Sequence 9697, Ap
c 24	67	3.1	6036	17	US-10-437-963-44135	Sequence 44135, A
c 25	65.4	3.0	5092	17	US-10-437-963-10162	Sequence 10162, A
c 26	64.8	3.0	1731	17	US-10-437-963-9134	Sequence 9134, Ap
c 27	62.8	2.9	2417	15	US-10-334-703-3	Sequence 3, Appli
c 28	62.8	2.9	2801	16	US-10-396-122-3	Sequence 3, Appli
c 29	62.8	2.9	2801	13	US-10-424-599-108650	Sequence 108650, Ap
c 30	62.8	2.9	9399	16	US-10-396-122-93	Sequence 93, Appl
c 31	62.6	2.9	1777	17	US-10-437-963-34059	Sequence 34059, A
c 32	61.2	2.8	5371	17	US-10-437-963-9498	Sequence 9498, Ap
c 33	61.2	2.8	9114	17	US-10-437-963-10728	Sequence 10728, A
c 34	61	2.8	4107	17	US-10-437-963-9190	Sequence 9190, Ap
c 35	60.2	2.8	3026	17	US-10-437-963-44142	Sequence 44142, A
c 36	59.6	2.8	9072	16	US-10-396-122-87	Sequence 87, Appl
c 37	59	2.7	3463	17	US-10-437-963-10660	Sequence 10660, A
c 38	58	2.7	5928	17	US-10-437-963-44133	Sequence 44133, A
c 39	57.6	2.7	4503	17	US-10-437-963-8782	Sequence 8782, Ap
c 40	56.6	2.6	7628	13	US-10-424-599-108640	Sequence 108640, A
c 41	56	2.6	3127	17	US-10-437-963-44134	Sequence 44134, A
c 42	53.4	2.5	4382	17	US-10-437-963-44138	Sequence 44138, A
c 43	53.4	2.5	5856	17	US-10-437-963-23904	Sequence 23904, A
c 44	53.2	2.5	13744	17	US-10-437-963-11092	Sequence 11092, A
c 45	52.8	2.4	5148	17	US-10-437-963-34840	Sequence 34840, A

#### ALIGNMENTS

#### RESULT 1

US-09-892-635A-44  
; Sequence 44, Application US/09892635A  
; Publication No. US20030226175A1  
; GENERAL INFORMATION:  
; APPLICANT: May, Gregory D.  
; APPLICANT: Glendennen, Stephanie K.  
; APPLICANT: Mason, Hugh S.  
; APPLICANT: Lim, Miguel A. Gomez  
; APPLICANT: Arntzen, Charles J.  
; TITLE OF INVENTION: DNA Regulatory Elements Associated with Fruit Development  
; FILE REFERENCE: 031998-007  
; CURRENT APPLICATION NUMBER: US/09/892, 635A  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 09/160,351  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: US 60/060,062  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 2156  
; TYPE: DNA  
; ORGANISM: Musa acuminata  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 879  
; OTHER INFORMATION: n = A,T,C or G  
US-09-892-635A-44

Query Match 99.9%; Score 2154.6; DB 13; Length 2156;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2156; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTGTATAGTTCAAGTTAGAAAAA 60  
|||||  
Db 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTGTATAGTTCAAGTTAGAAAAA 60  
|||||

Qy	61	TCTTTACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAA	CGTGTCATCATGCTCCTCGA	120
Db	61			
Qy	121	TGGACTCACCTGTTGTTTCAATTCGGAAGAAGTTCGAAAGAGTGCATAAGAAATATTGATTTGG	180	
Db	121			
Qy	181	ATTCTTTCACTCGGTGGTGCCTTCATGAGTGCACCTCAAGAGTCCCTCCAATATCAAAG	240	
Db	181			
Qy	241	CCGAATCACAAATTGAATGTGATTCGAATTCATTTTGTCTAATGACAAAACAGGCCAT	300	
Db	241			
Qy	301	TCATAGCCTTTGTGTTTAAAGCAAAAACATTTCTTCCGATTCATGCCATTCCGTCATCG	360	
Db	301			
Qy	361	GAAGAGAAAAATTTTGAATCCATTTTCGACAATAGACCAAAGCTCGAAATCCATCGAAA	420	
Db	361			
Qy	421	TGAGGAAGATCCTCATATGAGTTTTCOAATACATGTAATTCGACTCATTTAAACATAGGTTG	480	
Db	421			
Qy	481	GATGTGTAAATAAGATGACCCCTCATGCSCTATCTCTTTGGGTATTAACCAAATATGAGA	540	
Db	481			
Qy	541	GTGAGCCTTGCTCTGATACCAATTTGTTAGATCAGAGTGGCACTAAGAGAGGGGGGAGT	600	
Db	541			
Qy	601	GAATPAGTGCAGTGGATTA AAACTTTAAGTTTTAAAAATGAA	660	
Db	601			
Qy	661	TTTTGTTTTAATAGTAACTTTGAGTAGATGA AACCAAAGTTAACAGTAGTGTAAATAAC	720	
Db	661			
Qy	721	AATTTCCGGAAAAGTAAAGACTCA CACATTCAAAGAA CATACCAATTTAAAGTGGTTCGGT	780	
Db	721			
Qy	781	CAAAATGACCTPATCOACTTTGTGGAAGCCTTCTTTCGAAGAGGTCCCAA	840	
Db	781			
Qy	841	AAATCACTTTGAAGGGGAAGCAAAATACCTCTCTTACNACCTTTTACAATGGTTCATAC	900	
Db	841			
Qy	901	TCTTTACAAATTTTCAACGAGAAAGAGAGTGAACATGCAAGCAATTGAAAAACAAGACT	960	
Db	901			
Qy	961	TGCTTAAAGACTTTGCTTAAGGCTTTTTTTCTCAATCTAATTCGTTCTCAAAAGTTGATCTCT	1020	
Db	961			
Qy	1021	CTGCTGAGAAATGAGGGGTATTTATAGACCCCAGAGAGATTTAAATTTGGGCTCCAAATT	1080	
Db	1021			
Qy	1081	TCGAATGCTTTTGGGTTCCGAGGTTGCGGGTGCACCGGCTGTCACTGTTTGACACTGG	1140	
Db	1081			

Qy	1141	ACAGTGTAC	TAGCGGTG	CCACCGCCG	GAACCTCT	CTCGGGTGT	TGGCGCGT	GCACCGGCTAG	1201
Db	1141	ACAGTGTACT	TAGCGGTG	CCACCGCCG	GAACCTCT	CTCGGGTGT	TGGCGCGT	GCACCGGCTAG	1201
Qy	1201	ACTTTTTCAG	CTCACGTG	TTGGATTCC	CAAACTTGA	CCCAAAC	CAGTCCGAA	CTCGGTCC	1260
Db	1201	ACTTTTTCAG	CTCACGTG	TTGGATTCC	CAAACTTGA	CCCAAAC	CAGTCCGAA	CTCGGTCC	1260
Qy	1261	AATTGACCCG	TAAACCGGAT	TATAGGATTAA	CCCTTAA	CTTAAC	CCCTTAATATATG	CAAA	1320
Db	1261	AATTGACCCG	TAAACCGGAT	TATAGGATTAA	CCCTTAA	CTTAAC	CCCTTAATATATG	CAAA	1320
Qy	1321	CTAGCGAAG	CTGAAATATAG	TCCTAGCAG	TTTTTAA	CCGGCAAC	GTCCGACTTCTT		1380
Db	1321	CTAGCGAAG	CTGAAATATAG	TCCTAGCAG	TTTTTAA	CCGGCAAC	GTCCGACTTCTT		1380
Qy	1381	CCGGCGATCT	TTCCGGCAG	ACTTCGTATAT	ACCTTTGG	ATTCTT	TAGCGGAC	CTCCTAGT	1440
Db	1381	CCGGCGATCT	TTCCGGCAG	ACTTCGTATAT	ACCTTTGG	ATTCTT	TAGCGGAC	CTCCTAGT	1440
Qy	1441	AGGGTCCCG	ATCTTGTGG	CAGTTTACG	CGAGTTCG	GAACTTCT	CGGTGATCT	CCGCAA	1500
Db	1441	AGGGTCCCG	ATCTTGTGG	CAGTTTACG	CGAGTTCG	GAACTTCT	CGGTGATCT	CCGCAA	1500
Qy	1501	ACCGCGATG	ATCTTTCG	CAGACTTTCG	AAACTTCG	AAAGTCC	CGAATCTTCT		1560
Db	1501	ACCGCGATG	ATCTTTCG	CAGACTTTCG	AAACTTCG	AAAGTCC	CGAATCTTCT		1560
Qy	1561	GGTGTGGT	TCGACAG	CACTCTA	ACGAACTT	CGGACTC	CTCGAATG	CTCCAGACTTG	1620
Db	1561	GGTGTGGT	TCGACAG	CACTCTA	ACGAACTT	CGGACTC	CTCGAATG	CTCCAGACTTG	1620
Qy	1621	ACTCCGGT	AGGCTTG	CTTTATAT	TTTT	CAGGCTAT	CTAGTAA	TCCTACACTTAACTC	1680
Db	1621	ACTCCGGT	AGGCTTG	CTTTATAT	TTTT	CAGGCTAT	CTAGTAA	TCCTACACTTAACTC	1680
Qy	1681	AATAATATG	ATAGATTAA	TAAACCCAT	CAATTTG	ATTCTCAT	CATCAAA	ATTTCGACATTC	1740
Db	1681	AATAATATG	ATAGATTAA	TAAACCCAT	CAATTTG	ATTCTCAT	CATCAAA	ATTTCGACATTC	1740
Qy	1741	AACAAACAT	CCGTACTCA	ATAACCCAT	CAAGGCTAT	AGTTAC	GTGACTATCT	ACTGTGATC	1800
Db	1741	AACAAACAT	CCGTACTCA	ATAACCCAT	CAAGGCTAT	AGTTAC	GTGACTATCT	ACTGTGATC	1800
Qy	1801	CGTACGTGA	AGTTAG	CGAGTCAT	GATCCAG	GTGTCA	CTATTATG	CGCAACACGTATC	1860
Db	1801	CGTACGTGA	AGTTAG	CGAGTCAT	GATCCAG	GTGTCA	CTATTATG	CGCAACACGTATC	1860
Qy	1861	CTTTATCCAA	ATCCAGTCTT	CTCAACTCTT	CTAG	CCCTAC	CCGCTCTCTTTTTTTT	ATTACTT	1920
Db	1861	CTTTATCCAA	ATCCAGTCTT	CTCAACTCTT	CTAG	CCCTAC	CCGCTCTCTTTTTTTT	ATTACTT	1920
Qy	1921	TTGAAAGAA	TTCAATCA	AAACAGAT	CAAAATTA	ACACG	GTGACACTGT	GACATGCTA	1980
Db	1921	TTGAAAGAA	TTCAATCA	AAACAGAT	CAAAATTA	ACACG	GTGACACTGT	GACATGCTA	1980
Qy	1981	GTCTCTCGA	AGACATTA	TTTCGG	CGATPCCA	CAGACGT	CGTCAG	CTTTCATCACC	2040
Db	1981	GTCTCTCGA	AGACATTA	TTTCGG	CGATPCCA	CAGACGT	CGTCAG	CTTTCATCACC	2040
Qy	2041	TCCTACAT	ACATGTCC	ATGGCTTT	TGTTG	ATGACAG	ACACACAG	CTTGCTTTGGT	2100
Db	2041	TCCTACAT	ACATGTCC	ATGGCTTT	TGTTG	ATGACAG	ACACACAG	CTTGCTTTGGT	2100
Qy	2101	TGTGCCCT	TAAACAG	AGAGAGAG	AGACAC	CGATAG	CCCTCT	CAITCACTATGG	2156
Db	2101	TGTGCCCT	TAAACAG	AGAGAGAG	AGACAC	CGATAG	CCCTCT	CAITCACTATGG	2156



GENERAL INFORMATION:  
; APPLICANT: May, Gregory D.  
; APPLICANT: Clendennen, Stephanie K.  
; APPLICANT: Mason, Hugh S.  
; APPLICANT: Lim, Miguel A. Gomez  
; APPLICANT: Aritzen, Charles J.  
; TITLE OF INVENTION: DNA Regulatory Elements Associated with Fruit Development  
; FILE REFERENCE: 031998-007  
; CURRENT APPLICATION NUMBER: US/09/892,635A  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 09/160,351  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: US 60/060,062  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 2160  
; TYPE: DNA  
; ORGANISM: Musa acuminata  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 883  
; OTHER INFORMATION: n = A,T,C or G  
; US-09-892-635A-45

Query Match 99.2%; Score 2139; DB 13; Length 2160;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2155; Conservative 0; Mismatches 1; Indels 4; Gaps 1;  
QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATATAGTTTCAAAAGTTAGAAAA 60  
Db 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATATAGTTTCAAAAGTTAGAAAA 60  
QY 61 TCATTACCAAGAGCTTTGAGTCCATTGATGACATCCCGTGAACCGGTGTACATGCTCCGA 120  
Db 61 TCATTACCAAGAGCTTTGAGTCCATTGATGACATCCCGTGAACCGGTGTACATGCTCCGA 120  
QY 121 TGGACTCAGTGGTTTCATTCGGAAGAGTTTGAAGAGTGCATAGAAATATTGATTTGG 180  
Db 121 TGGACTCAGTGGTTTCATTCGGAAGAGTTTGAAGAGTGCATAGAAATATTGATTTGG 180  
QY 181 ATTCTTTCACTCGGTTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAG 240  
Db 181 ATTCTTTCACTCGGTTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAG 240  
QY 241 CCGAATCACAATTTGAATGTGAATTTGATTTGCTTAATGCAACAAACAGGGCAT 300  
Db 241 CCGAATCACAATTTGAATGTGAATTTGATTTGCTTAATGCAACAAACAGGGCAT 300  
QY 301 TCATAGCCTTTGTTTAAAGCAAAACATCTTCCGATTCATCCCATTCGCTCATCG 360  
Db 301 TCATAGCCTTTGTTTAAAGCAAAACATCTTCCGATTCATCCCATTCGCTCATCG 360  
QY 361 GAAGAGAAAAATTTTGAATCCATTTTCGACAATAGACCAAGCTCGAAATC---CATG 416  
Db 361 GAAGAGAAAAATTTTGAATCCATTTTCGACAATAGACCAAGCTCGAAATCCATGATG 420  
QY 417 GAAATGAGGAAGATCTTCATATAGATTTTCAATATACATGTAATTCGACTCATTAACATA 476  
Db 421 GAAATGAGGAAGATCTTCATATAGATTTTCCAAATACATGTAATTCGACTCATTAACATA 480  
QY 477 GGTGGATGTGAATGAATGACCTCATGCTATCTCTCGGTATTAACCAATAT 536  
Db 481 GGTGGATGTGAATGAATGACCTCATGCTATCTCTCGGTATTAACCAATAT 540  
QY 537 GAGAGTGAGCTTGTCTGATACCAATTTGTAGATCAGAGTGGCACTTAAGAGAGGGGG 596  
Db 541 GAGAGTGAGCTTGTCTGATACCAATTTGTAGATCAGAGTGGCACTTAAGAGAGGGGG 600  
QY 597 GAGTGAATAGTCAGTGGATTAACCTTATAAGTTTAAAAATGAATTCGTAATAACGAG 656  
Db 601 GAGTGAATAGTCAGTGGATTAACCTTATAAGTTTAAAAATGAATTCGTAATAACGAG 660

QY 657 AAGATTTTCGTTTATAGTAGTAACTTAGATGATGAAAAACCAAAAGTTAACAGTAGTGAAA 716  
Db 661 AAGATTTTCGTTTATAGTAGTAACTTAGATGATGAAAAACCAAAAGTTAACAGTAGTGAAA 720  
QY 717 TAACAAATTTTCGGGAAAGTAGAAGTCAACATTCAGAGAAACATACCAATTTAAAGTGGTT 776  
Db 721 TAACAAATTTTCGGGAAAGTAGAAGTCAACATTCAGAGAAACATACCAATTTAAAGTGGTT 780  
QY 777 CGGTCAAAATGACCTACATCCACTTGTGAAGCCCTTCTTGAAGAGGCTCCCAACTCCAC 836  
Db 781 CGGTCAAAATGACCTACATCCACTTGTGAAGCCCTTCTTGAAGAGGCTCCCAACTCCAC 840  
QY 837 TAGCAATCACTTTGAAGGGAAGGACAAATACCTCTCTACNACCTTTTACAATGGTTC 896  
Db 841 TAGCAATCACTTTGAAGGGAAGGACAAATACCTCTCTTACNACCTTTTACAATGGTTC 900  
QY 897 ATACTCTTACAAATTTTCAACGAGAAAGAGGAGTGAACATGCAAGCAATTTGAAAACAA 956  
Db 901 ATACTCTTACAAATTTTCAACGAGAAAGAGGAGTGAACATGCAAGCAATTTGAAAACAA 960  
QY 957 GACTTGCTAAAGACTTTTGCTAAAGCTTTTTTCTCAATCTATTGCTTCTCAAAAGTTGTA 1016  
Db 961 GACTTGCTAAAGACTTTTGCTAAAGCTTTTTTCTCAATCTATTGCTTCTCAAAAGTTGTA 1020  
QY 1017 TTCTCTGCTGAGATTTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCA 1076  
Db 1021 TTCTCTGCTGAGATTTGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCA 1080  
QY 1077 AATTTTCAAAATGCTTTCCGAGTTTCCGAGTTTCCGAGTGCACCGCTCTCAGTGTTCGACA 1136  
Db 1081 AATTTTCAAAATGCTTTCCGAGTTTCCGAGTTTCCGAGTGCACCGCTCTCAGTGTTCGACA 1140  
QY 1137 CTGACAGAGTGTATAGCGGTGCCACCGCCGAGCTCTCGGGTGTGGGGCGTGCACCGC 1196  
Db 1141 CTGACAGAGTGTATAGCGGTGCCACCGCCGAGCTCTCGGGTGTGGGGCGTGCACCGC 1200  
QY 1197 CTAGACTTTTTCAGTCTCAGTGTGGATTTCAAACTTGACCCAAACAGTCCGACTCGG 1256  
Db 1201 CTAGACTTTTTCAGTCTCAGTGTGGATTTCAAACTTGACCCAAACAGTCCGACTCGG 1260  
QY 1257 GTCCAAATGACCGGTAAACCGGATTTAGAGTAAACCTTAATCTTAACCCCTAATATATG 1316  
Db 1261 GTCCAAATGACCGGTAAACCGGATTTAGAGTAAACCTTAATCTTAACCCCTAATATATG 1320  
QY 1317 CAAACTACGCACTGAAAAATATAGTCTTAAGCAAGTTTTTAAACCGGCAACGTCGAGTCT 1376  
Db 1321 CAAACTACGCACTGAAAAATATAGTCTTAAGCAAGTTTTTAAACCGGCAACGTCGAGTCT 1380  
QY 1377 TCTTCGGCGATCTTTTCGGCAGACTTCTGATATACCTTTGGATTTCTTCTAGCGGACTCC 1436  
Db 1381 TCTTCGGCGATCTTTTCGGCAGACTTCTGATATACCTTTGGATTTCTTCTAGCGGACTCC 1440  
QY 1437 TAGTAGGTTCCCGATCTTGTGGCGAGTTTAGCGAGTAGCGGAACCTTCTCGGTGATCTCC 1496  
Db 1441 TAGTAGGTTCCCGATCTTGTGGCGAGTTTAGCGAGTAGCGGAACCTTCTCGGTGATCTCC 1500  
QY 1497 GCAAAACCGCGATGATCTCTTCGGCAGACTTTCGAAAACTTCGACAAGTCCCGGATTTCT 1556  
Db 1501 GCAAAACCGCGATGATCTCTTCGGCAGACTTTCGAAAACTTCGACAAGTCCCGGATTTCT 1560  
QY 1557 TCTTCGGTTGGTTCGAGCAGATCTCTAAACGAACTTCGAACTCTTGAATGTCATCGAA 1616  
Db 1561 TCTTCGGTTGGTTCGAGCAGATCTCTAAACGAACTTCGAACTCTTGAATGTCATCGAA 1620  
QY 1617 CTTGACTCCGGTAGGCTTGTCTTATATTTTTCAGGCTATCATAGTAACTCTACATCTTA 1676  
Db 1621 CTTGACTCCGGTAGGCTTGTCTTATATTTTTCAGGCTATCATAGTAACTCTACATCTTA 1680  
QY 1677 ACTCAATAATATGATTAATTAATTAACCCCAATTAATTTTCAATCATCAAAATTCGAC 1736  
Db 1681 ACTCAATAATATGATTAATTAATTAACCCCAATTAATTTTCAATCATCAAAATTCGAC 1740

```
QY 1737 ATTCAACAAACATCGTACTCAATAAACCATCAGGCTATAGTTAGTCACTATCTACTGT 1796
|
|
|
Db 1741 ATTCAACAAACATCCGTACTCAATAAACCCATCAGGCTATAGTTAGTCACTATCTACTGT 1800
|
|
|
QY 1797 GATCCGTACGTTAGGAGTTCATGATCCAGGTCGTGTCACTTATGGCCGAAACACG 1856
|
|
|
Db 1801 GATCCGTACGTTAGGAGTTCATGATCCAGGTCGTGTCACTTATGGCCGAAACACG 1860
|
|
|
QY 1857 TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGGCTTACCGGTCCTTTTATTT 1916
|
|
|
Db 1861 TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGGCTTACCGGTCCTTTTATTT 1920
|
|
|
QY 1917 ACTTTTGAAGAAATTCAAATCAAAACAGATACAAATAAACAACGCTGAGACATGTGACAT 1976
|
|
|
Db 1921 ACTTTTGAAGAAATTCAAATCAAAACAGATACAAATAAACAACGCTGAGACATGTGACAT 1980
|
|
|
QY 1977 GCTAGTCTCTCGAAAGCAATTAATTCGCGCATCCACAGACGTCGTCAAGCTTCAACCCAC 2036
|
|
|
Db 1981 GCTAGTCTCTCGAAAGCAATTAATTCGCGCATCCACAGACGTCGTCAAGCTTCAACCCAC 2040
|
|
|
QY 2037 TTTTTCCTACATACATGTCGATGCGCTTTGTTGTGATGACAGACACCAACGCTTGCCTT 2096
|
|
|
Db 2041 TTTTTCCTACATACATGTCGATGCGCTTTGTTGTGATGACAGACACCAACGCTTGCCTT 2100
|
|
|
QY 2097 TGGTTGTGCTTAAACAGAGAGAGAGAGACAGACCGATAGCCTCTCATTCATATGG 2156
|
|
|
Db 2101 TGGTTGTGCTTAAACAGAGAGAGAGAGACAGACCGATAGCCTCTCATTCACCATGG 2160
|
|
|
RESULT 3
US-09-892-635a-27
; Sequence 27, Application US/09892635A
; Publication No. US20030226175A1
; GENERAL INFORMATION:
; APPLICANT: May, Gregory D.
; APPLICANT: Clendenen, Stephanie K.
; APPLICANT: Mason, Hugh S.
; APPLICANT: Lim, Miguel A. Gomez
; APPLICANT: Arntzen, Charles J.
; TITLE OF INVENTION: DNA Regulatory Elements Associated with Fruit Development
; CURRENT APPLICATION NUMBER: US/09/892,635A
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 09/160,351
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: US 60/060,062
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 4924
; TYPE: DNA
; ORGANISM: Musa acuminata
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 879..3691, 4119
; OTHER INFORMATION: n = A,T,C or G
US-09-892-635a-27
Query Match 98.6%; Score 2125.6; DB 13; Length 4924;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2152; Conservative 0; Mismatches 0; Indels 5; Gaps 2;
QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTAGTTATTAAGTTCAAGTTAGAAAA 60
Db 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTAGTTATTAAGTTCAAGTTAGAAAA 60
QY 61 TCTTTTACCAAGAGCTTTCAGTCCATTGATGACATCCGTGACCGGTGATGTCCTCCGA 120
Db 61 TCTTTTACCAAGAGCTTTCAGTCCATTGATGACATCCGTGACCGGTGATGTCCTCCGA 120
QY 121 TGGACTCACTTGGTTTCATTCGAAAAAGTTCCGAAAGAGTGATCAATAAGATTTTGG 180
|
|
|
```

```
Db 121 TGGACTCACTTGGTTTCATTCGAAAAAGTTCCGAAAGAGTGATCAATAAGATTTTGG 180
QY 181 ATTCTTTCACCTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAG 240
|
|
|
Db 181 ATTCTTTCACCTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAG 240
|
|
|
QY 241 CCGAATCACAAATTTGAAATGTGAATTCATTTTGTCTAAATGTCACAAAACAGGGCAT 300
|
|
|
Db 241 CCGAATCACAAATTTGAAATGTGAATTCATTTTGTCTAAATGTCACAAAACAGGGCAT 300
|
|
|
QY 301 TCATAGCCTTTGTCTTTAAAGCAAAAAACATCTCTCCGATTCATCCCATTCGTCATCG 360
|
|
|
Db 301 TCATAGCCTTTGTCTTTAAAGCAAAAAACATCTCTCCGATTCATCCCATTCGTCATCG 360
|
|
|
QY 361 GAAGAGAAAAATTTTGAATTCATTTTGAATTCATTTTGAATTCATTTTGAATTCATTTTGA 420
|
|
|
Db 361 GAAGAGAAAAATTTTGAATTCATTTTGAATTCATTTTGAATTCATTTTGAATTCATTTTGA 420
|
|
|
QY 421 TGAGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTTAAACATAGTG 480
|
|
|
Db 421 TGAGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTTAAACATAGTG 480
|
|
|
QY 481 GATGTGTAATGAATGACCCCTCATGCSCTATCTCTCTTGGGTATTAACCAAAATATGAGA 540
|
|
|
Db 481 GATGTGTAATGAATGACCCCTCATGCSCTATCTCTCTTGGGTATTAACCAAAATATGAGA 540
|
|
|
QY 541 GTGAGCCTTGTCTGATACCAATTTGTTAGGATCAGAGTGGCCTAAAGAGAGGGGGAGT 600
|
|
|
Db 541 GTGAGCCTTGTCTGATACCAATTTGTTAGGATCAGAGTGGCCTAAAGAGAGGGGGAGT 600
|
|
|
QY 601 GAATTAGTGACGTGATTAATAAATTTAAATGTTAAATGTAATTCGTAATACGAGAAGA 660
|
|
|
Db 601 GAATTAGTGACGTGATTAATAAATTTAAATGTTAAATGTAATTCGTAATACGAGAAGA 660
|
|
|
QY 661 TTTCCGTTTAAATAGTAATCCTGAGTAGATGAAAAACAAAAGTTAACAGTAGTGTAAATAC 720
|
|
|
Db 661 TTTCCGTTTAAATAGTAATCCTGAGTAGATGAAAAACAAAAGTTAACAGTAGTGTAAATAC 720
|
|
|
QY 721 AATTTCCGGAAGTAAGAACTCACACATTCAAGGAACATACCAATTTAAAGTGTTCGGT 780
|
|
|
Db 721 AATTTCCGGAAGTAAGAACTCACACATTCAAGGAACATACCAATTTAAAGTGTTCGGT 780
|
|
|
QY 781 CAAATGACCTACATCCACTTGTGAAAGCCTTCTCGAAGAGGCTCCCAACTTCCACTAGC 840
|
|
|
Db 781 CAAATGACCTACATCCACTTGTGAAAGCCTTCTCGAAGAGGCTCCCAACTTCCACTAGC 840
|
|
|
QY 841 AAATCACTTTGAAGGGGGAAGACAAATACCTCTCTTACNAGCTTTTACAAATGGTTCATAC 900
|
|
|
Db 841 AAATCACTTTGAAGGGGGAAGACAAATACCTCTCTTACNAGCTTTTACAAATGGTTCATAC 900
|
|
|
QY 901 TCTTACAAATTTTCAACGAGAAAGAGGAGTGACATGCAAGCAATTTGAAACCAAGACT 960
|
|
|
Db 901 TCTTACAAATTTTCAACGAGAAAGAGGAGTGACATGCAAGCAATTTGAAACCAAGACT 960
|
|
|
QY 961 TGCTAAAGACTTTGCTAAGGCTTTTTTTCTCAATCTATTGCTTCTCAAAAGTGTATTTCT 1020
|
|
|
Db 961 TGCTAAAGACTTTGCTAAGGCTTTTTTTCTCAATCTATTGCTTCTCAAAAGTGTATTTCT 1020
|
|
|
QY 1021 CTGCTGAGAAATGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080
|
|
|
Db 1021 CTGCTGAGAAATGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCAAAT 1080
|
|
|
QY 1081 TCGAATGCTTCTGGGTTCCCGAGGTTCCCGGTCGCCCGCTGTGAGTGTGTGACACTGG 1140
|
|
|
Db 1081 TCGAATGCTTCTGGGTTCCCGAGGTTCCCGGTCGCCCGCTGTGAGTGTGTGACACTGG 1140
|
|
|
QY 1141 ACAGTGTACTAGCGGTGCCCGCGGACCTCTCGGGTGTGGGGGTGTGCGGCTGACCCCTAG 1200
|
|
|
Db 1141 ACAGTGTACTAGCGGTGCCCGCGGACCTCTCGGGTGTGGGGGTGTGCGGCTGACCCCTAG 1200
|
|
|
QY 1201 ACTTTTTCAGCTCACTGTTGGATTCCTCAAACTTGACCCCAACCACTGCGGCTCGGTC 1260
|
|
|
Db 1201 ACTTTTTCAGCTCACTGTTGGATTCCTCAAACTTGACCCCAACCACTGCGGCTCGGTC 1260
|
|
|
```

```
1261 AATTGACCGTAACCGGATTATAGATTAAACCCCTTAATCCTAAACCCCTAATATATGCAAA 1320
1261 AATTGACCGTAACCGGATTATAGATTAAACCCCTTAATCCTAAACCCCTAATATATGCAAA 1320
1321 CTACGGCACTGAAATATAGCTTAAAGCAAGTTTAAACCGGCAAAACGTCGAGTCTTCTT 1380
1321 CTACGGCACTGAAATATAGCTTAAAGCAAGTTTAAACCGGCAAAACGTCGAGTCTTCTT 1380
1381 CCGGGATCTTTCCGAGACTTCTGATATACCTTTTGGATTCTTCTAGCGACTCCCTAGT 1440
1381 CCGGGATCTTTCCGAGACTTCTGATATACCTTTTGGATTCTTCTAGCGACTCCCTAGT 1440
1441 AGGTCGCGATCTTTGCGGAGTTTAGCGAGTAGCGAAGCTTCTCGGTGATCTCCGCAA 1500
1441 AGGTCGCGATCTTTGCGGAGTTTAGCGAGTAGCGAAGCTTCTCGGTGATCTCCGCAA 1500
1501 ACCCGGATGATCTTTCCGAGACTTTTCGAAACTTCGCAAGTCCCGGATTTCTTCTC 1560
1501 ACCCGGATGATCTTTCCGAGACTTTTCGAAACTTCGCAAGTCCCGGATTTCTTCTC 1560
1561 GGTGGTTCGACAGATCTCTAACGAACTTCGCAAGTCCCGGATTTCTTCTC 1620
1561 GGTGGTTCGACAGATCTCTAACGAACTTCGCAAGTCCCGGATTTCTTCTC 1620
1621 ACTCGGTAGCTTCTTATATTTTCAAGGCTATCATAGTTAATCCTACATTAATCTC 1680
1621 ACTCGGTAGCTTCTTATATTTTCAAGGCTATCATAGTTAATCCTACATTAATCTC 1680
1681 AATAATATGATTAGATTAAATTAACCCATCAATTAATTAATTAATTAATTAATTAATTA 1740
1681 AATAATATGATTAGATTAAATTAACCCATCAATTAATTAATTAATTAATTAATTAATTA 1740
1741 AACAAATCGTACTCAATAACCCATCAATTAATTAATTAATTAATTAATTAATTAATTA 1800
1741 AACAAATCGTACTCAATAACCCATCAATTAATTAATTAATTAATTAATTAATTAATTA 1800
1801 CGTACGTAAGTATAGCGATCATCATCAGGTCGTCACCTTATGCGGCAACACGTCATC 1860
1801 CGTACGTAAGTATAGCGATCATCATCAGGTCGTCACCTTATGCGGCAACACGTCATC 1860
1861 CTTATCCAAATCCAGTCTTCTCAACTCTTACGCTTACCGCTCTCTTTTATTAATCTT 1920
1861 CTTATCCAAATCCAGTCTTCTCAACTCTTACGCTTACCGCTCTCTTTTATTAATCTT 1920
1921 TTGAAAGAAATCAAAATCAAAACAGATACAAATTAACACGCTGAGACACTGTGACATGCTA 1980
1921 TTGAAAGAAATCAAAATCAAAACAGATACAAATTAACACGCTGAGACACTGTGACATGCTA 1980
1981 GTCTCTGAAAGCAATTAATTCGCGCATCCACAGAGTCGTCAGCTTTCATCACCCTTTT 2040
1981 GTCTCTGAAAGCAATTAATTCGCGCATCCACAGAGTCGTCAGCTTTCATCACCCTTTT 2040
2041 TCCTACAT-ACCATGTCGATGGCTTTGTTGATGACAGACACCAACAGCTTGCCTTTGG 2099
2041 TCCTACATACCATGTCGATGGCTTTGTTGATGACAGACACCAACAGCTTGCCTTTGG 2100
2100 TTGTGCTTACAGAGAGAGAGAGACAGACAGCGATGCTCTCTCATTCACATG 2156
2101 TTGTGCTTAC-AGAGAGAGAGAGACAGACAGCGATGCTCTCTCATTCACATG 2153
```

## RESULT 4

US-10-298-122-3

```
; Sequence 3, Application US/10298122
; Publication No. US200302212141
; GENERAL INFORMATION:
; APPLICANT: Gmitter, Frederick G
; APPLICANT: Deng, Zhanao
; APPLICANT: Zhang, Hongbin
; TITLE OF INVENTION: Citrus Tristeza Virus and Methods of Use
; FILE REFERENCE: 5853-220
; CURRENT APPLICATION NUMBER: US/10/298,122
```

```
; CURRENT FILING DATE: 2002-11-15
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
```

LENGTH: 78333

TYPE: DNA

ORGANISM: artificial

FEATURE:

```
; OTHER INFORMATION: Poncirus trifoliata and citrus grandis hybrid
US-10-298-122-3
```

## Query Match

Best Local Similarity 5.9%; Score 127.6; DB 16; Length 78333;

Matches 298; Conservative 0; Mismatches 199; Indels 31; Gaps 2;

```
Qy 6 CCACTTTTGGAGTATGATCTTAAATTTTATAGTTTAAAGTTTCAAGTTTGAAGAAATCTTT 65
Db 47226 CCACTCTTAGGTAGATCTAATGATTTTCTTGACCTTTTCGCTATTGAAAGTCTT 47285
Qy 66 ACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAAC-GGTGTACATGCTCCGATGGA 124
Db 47286 TCCTAGGGCTCCTAGAGTGTTCATATGTCGTGAATCTAGTATACATAGAGTACACATT 47345
Qy 125 CTCACCTTGGTTTCATTCGGAAGAGTTCGAAAGAGTGCATBAAGATATATTGTTGGATTC 184
Db 47346 TTCATTTTGTTCATTTGAAACCAATTCGTATTTGCGAGTGTCTACTTATTTTAGATTC 47405
Qy 185 TTTCACTCGGTTGGTGCCTTCATGAGTACCTCAAGAGTCTCCAAATATCAAAAGCGGA 244
Db 47406 TTTCACTTGGTTCGTACCTTCATAGACACCTCAAGTTTGTGCCAAATCTCATTTAGCACT 47465
Qy 245 ATACAAATTTGAAATGATGATGATTTTGTCTAATGCAAAACAGGGCAATTCAT 304
Db 47466 TTCGCACTAGACACTCTATGAAACTCTTACTTATCTAGTGCACAGAACAGGCAATTCAT 47525
Qy 305 AGCTTTGTCTTTAAAGCAAAACATTTCTTCGATTCATCCATTCGCTCATCGGAAG 364
Db 47526 GGCTTGGAAATTTAGAGAGCTTTTCTTATCGAATTCATCCAAATCCCGTGAAGTTT 47585
Qy 365 AGAAATTTTGAAT-----CCATTTTCGCAAT 394
Db 47586 TGGATATCTTCTCTACTTCAATTTTAGTCAAAGGCATGAAGGCCCATCATTAACAAT 47645
Qy 395 AGACCAAGCTCGAATCCATGGAATAGAGAGATCCTCATATGAGTTTTCCAATACAT 454
Db 47646 TTCACCAATCTGTAATCTTAAGGCTTGAAGTAAATTCATCTAGTTTTCCAATATGG 47705
Qy 455 GTAATTCGACTCATTAACACATAGTGGATGTAATGAATGACCTC 502
Db 47706 GTAATCATACCATCAAGTAAGGGGTCTAGTGGGACTGCTTC 47753
```

## RESULT 5

US-10-298-122-2

Sequence 2, Application US/10298122

Publication No. US200302212141

GENERAL INFORMATION:

APPLICANT: Gmitter, Frederick G

APPLICANT: Deng, Zhanao

APPLICANT: Zhang, Hongbin

TITLE OF INVENTION: CITRUS TRISTEZA VIRUS AND METHODS OF USE

FILE REFERENCE: 5853-220

CURRENT APPLICATION NUMBER: US/10/298,122

CURRENT FILING DATE: 2002-11-15

NUMBER OF SEQ ID NOS: 19

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 58965

TYPE: DNA

ORGANISM: artificial

FEATURE:

OTHER INFORMATION: Poncirus trifoliata and Citrus grandis hybrid

```
; NAME/KEY: misc feature
; LOCATION: (2145)..(2320)
; OTHER INFORMATION: n denotes unsequenced nucleotides
US-10-298-122-2

Query Match 4.3%; Score 92; DB 16; Length 58965;
Best Local Similarity 57.8%; Pred. No. 1.2e-12;
Matches 186; Conservative 0; Mismatches 130; Indels 6; Gaps 1;

Qy 181 ATTCTTTTCACTCGGTGGTGGCTTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAG 240
Db 38730 ATTCTCTCTCTTGTGTAGTCTCTCATACACAAATTTTGAGTTTCTCCAAATTCATAAG 38789

Qy 241 CCGAATCACAATGAATGATGAATTCATTTTGTCTAATGCACAAAACAGGGCAT 300
Db 38790 CATTAGAACAAATTTGAAACTCTATGAAATTCCTTTTATCTAAGGCACAAAATAAGCAT 38849

Qy 301 TCATAGCCTTTGTGTTTAAAGCAAAACATTCCTTCGGATTCTATCTAATCTCACTTAAGCATTCITT 38909
Db 38950 TCATAGCTTTGGAATTTAAACACATATTTTCTTATCTAATCTCACTTAAGCATTCITT 38909

Qy 361 GAAGAGAAAATTTTGAATCCATTTTCGACAAATAGACCAAGGCTCGAAATCCCATGGAAA 420
Db 38910 TAGGA-----ATGAGTAAATCATCACAAACAACTTCCCAAATTTTCAATATCTAAAGCTT 38963

Qy 421 TGAGGAAGATCTCTATATGAGTTTTCATATCATATGATTTTCGACTCATTAACATAGTGT 480
Db 38964 GTAAATAAATTCATCTAGTTTTCCTCAATAAGGATAATCATTTCCATCTAAATAATGGAG 39023

Qy 481 GATGTGTAATGAATGACCCCTC 502
Db 39024 GTCITGTGATAGATTGCTCTTC 39045

RESULT 6
US-10-424-599-108338
; Sequence 108338, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 108338
; LENGTH: 2760
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_68844C.1
US-10-424-599-108338

Query Match 3.8%; Score 81.4; DB 13; Length 2760;
Best Local Similarity 53.8%; Pred. No. 1.2e-10;
Matches 189; Conservative 0; Mismatches 161; Indels 1; Gaps 1;

Qy 1  GGATCCCACTTTTAGGATGATCTTAAATTTTAGTTTATAGTTCAAGTTAGAGTACAAAAA 60
Db 1829 GGTTCGCATCTCTACTTAAACATCTTAAACCTTGTATTATAGATCTCATATTTGAAAT 1888

Qy 61 TCTTTTACCAAGAGCTTTGAGTCCATGATGACATCCGTGAAACGGGTGACATGCTCCG- 119
Db 1889 TCTTTTGCCTAAGGCTGCTAGATGATTTACTATATGTGTAATCTCTTTTGCATACCTGCA 1948

Qy 120 ATGGACTCACTTGGTTTCAATTCGGAAGCTTCGAAGAGTGCATAGATATTTGATTG 179
Db 1949 ATATTTTTCATTTGCATTCATCTTAAATAATTCATCTCATGAGTTAGTGCATTTATCCTA 2008

180 GATTCCTTTTCACTCGGTGGTGGCTTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAA 239
Db 2009 GATCTTTTAAACATTTGATGTAGTTCCTTTCATGTGTTAATCGAAGAGTGTCCACATTTCTTA 2068

Qy 240 GCGAATCACAATTTGAAATGTGATGAATTCATTTTGTCTAATGCACAAAACAGGGCA 299
Db 2069 GCATCTTCAAAATTTGAAACCTCGAAATATTCATCCATTCCTAGGCGAGATGTTATTATG 2128

Qy 300 TTCATAGCCTTTGTGTTTAAAGCAAAACATTCCTTCCTCCGATTCATCCCAT 350
Db 2129 TTTTGGCTTTTAAATTTGATTTGATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2179

RESULT 7
US-09-966-881-48
; Sequence 48, Application US/09966881
; Patent No. US20020120960A1
; GENERAL INFORMATION:
; APPLICANT: Seymour, Graham
; Bird, Colin
; Medina-Suarez, Rosybel
; TITLE OF INVENTION: Genetic control Of Fruit Ripening
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zeneca Ag Products Inc.
; STREET: 1800 Concord Pike
; CITY: Wilmington
; STATE: DE
; COUNTRY: USA
; ZIP: 19850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/966,881
; FILING DATE: 28-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/242,860
; FILING DATE: 29-Mar-1999
; APPLICATION NUMBER: GB 9618862.8
; FILING DATE: 10-SEP-1996
; APPLICATION NUMBER: GB 9708366.1
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: PCT/GB97/02424
; FILING DATE: 08-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: SEE 50183/UST
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 886-1699
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: U-D86
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
US-09-966-881-48

Query Match 3.5%; Score 75; DB 9; Length 809;
Best Local Similarity 100.0%; Pred. No. 2.9e-09;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2082 CCAGAGCTTCCTTTGGTTGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2141
Db 9 CCACAAAGCTTGCCTTTGGTTGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68
```

QY 2142 CCTCATTCACATGCG 2156  
DB 69 CCTCATTCACATGCG 83

RESULT 8  
US-10-437-963-10782/c  
; Sequence 10782, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 10782  
; LENGTH: 5016  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1706C.1  
US-10-437-963-10782

Query Match 3.4%; Score 73.4; DB 17; Length 5016;  
Best Local Similarity 49.3%; Pred. No. 2.5e-08;  
Matches 251; Conservative 1; Mismatches 247; Indels 10; Gaps 2;

QY 4 TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAATCT 63  
DB 500 TCATACCTTCGCGTAGGGCTCTAAGCATCTTTTGGGCAACCTCAAGATCGGTAGTTT 441

QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTTGAAC-GGTGTACATGTCCTCCGATG 122  
DB 440 GCCCAAGGCCCTTGAGATCAATCACAATGACATTTGAGCTCCCATACATGCTATTCA 381

QY 123 GACTCATTGCTTTCATTCGGAAGTTTCGAAAGAGTGCATAAGAAATATTGATTTTGGAT 182  
DB 380 CTCTCATGAGGCAACATGGAGAATGTCATATTGGATCTTGAGAAATGAAGCTTGGCA 321

QY 183 TCCTTCACTCGGTGGTCCCTTCATGAGTCACTCAAGAGTCCCAAGATTAATTTGGAT 242  
DB 320 TCCTTGTACTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 261

QY 183 TCCTTCACTCGGTGGTCCCTTCATGAGTCACTCAAGAGTCCCAAGATTAATTTGGAT 242  
DB 320 TCCTTGTACTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 261

QY 243 GAATCAAAATTTGAAATTCATTTTCGACATAGCAACCAAGCTCGAATCCATGGAATG 302  
DB 320 TCCTTGTACTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 261

QY 243 GAATCAAAATTTGAAATTCATTTTCGACATAGCAACCAAGCTCGAATCCATGGAATG 302  
DB 320 TCCTTGTACTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 261

QY 303 ATAGCCTTTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCAATTCGCTCATCGGA 362  
DB 200 ATGGC-----TTGGGCATTGAGTTGGAGGTACGGGTGATCACTCTCCGTCAGGGT 150

QY 363 AGAGAAAATTTTGAATTCATTTTCGACATAGCAACCAAGCTCGAATCCATGGAATG 422  
DB 149 GTCCCGGTGATATCAAGCCCTACATCCAAATATCTCCAAATATGGAAGCTCATAGCTTTG 90

QY 423 AGAAGATCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAAACATAGGTGGA 482  
DB 89 AGGTAGTAGACATTTTAAATTTCCAAAGTGGAGTAGTTTGTGCCATTAAACATGGAAGCC 30

QY 483 TGTGTAATGAAATGACCCCTCATGCSCTAT 511  
DB 29 TTCCCTACATGGTTCACTCGTTGACAT 1

RESULT 10  
US-10-437-963-10117  
; Sequence 10117, Application US/10437963  
; Publication No. US20040123343A1

RESULT 9  
US-10-437-963-44132/c  
; Sequence 44132, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 44132  
; LENGTH: 5304  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_47222C.1  
US-10-437-963-44132

Query Match 3.4%; Score 73.4; DB 17; Length 5304;  
Best Local Similarity 49.3%; Pred. No. 2.6e-08;  
Matches 251; Conservative 1; Mismatches 247; Indels 10; Gaps 2;

QY 4 TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAATCT 63  
DB 500 TCATACCTTCGCGTAGGGCTCTAAGCATCTTTTGGGCAACCTCAAGATCGGTAGTTT 441

QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTTGAAC-GGTGTACATGTCCTCCGATG 122  
DB 440 GCCCAAGGCCCTTGAGATCAATCACAATGACATTTGAGCTCCCATACATGCTATTCA 381

QY 123 GACTCATTGCTTTCATTCGGAAGTTTCGAAAGAGTGCATAAGAAATATTGATTTTGGAT 182  
DB 380 CTCTCATGAGGCAACATGGAGAATGTCATATTGGATCTTGAGAAATGAAGCTTGGCA 321

QY 183 TCCTTCACTCGGTGGTCCCTTCATGAGTCACTCAAGAGTCCCAAGATTAATTTGGAT 242  
DB 320 TCCTTGTACTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 261

QY 243 GAATCAAAATTTGAAATTCATTTTCGACATAGCAACCAAGCTCGAATCCATGGAATG 302  
DB 260 GTCTCAAGTTGCTCACTCTATCGAATCTCTTCCGCACTCTCCGCACTCTCATATGCG 201

QY 303 ATAGCCTTTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCAATTCGCTCATCGGA 362  
DB 200 ATGGC-----TTGGGCATTGAGTTGGAGGTACGGGTGATCACTCTCCGTCAGGGT 150

QY 363 AGAGAAAATTTTGAATTCATTTTCGACATAGCAACCAAGCTCGAATCCATGGAATG 422  
DB 149 GTCCCGGTGATATCAAGCCCTACATCCAAATATCTCCAAATATGGAAGCTCATAGCTTTG 90

QY 423 AGAAGATCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAAACATAGGTGGA 482  
DB 89 AGGTAGTAGACATTTTAAATTTCCAAAGTGGAGTAGTTTGTGCCATTAAACATGGAAGCC 30

QY 483 TGTGTAATGAAATGACCCCTCATGCSCTAT 511  
DB 29 TTCCCTACATGGTTCACTCGTTGACAT 1





OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1690C.1  
US-10-437-963-10605

Query Match	3.3%;	Score 71.8;	DB 17;	Length 10369;
Best Local Similarity	49.1%;	Pred. No. 1.1e-07;		
Matches 250;	Conservative 1;	Mismatches 248;	Indels 10;	Gaps 2;

  

QY	4	TCCCAACTTTAGAAATGGATCTTAAAAATTTAGTTTATAAGTTCAAAGTTAGAAAAATCT	63
Db	9870	TCATCTTCTCCGGTAGGCTCTAAGCATCTTTTGGGCAACCTCAAGATCGGTGTAGTTT	9929
QY	64	TTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTGAAAC-GGTGTACATGTCTCCGATG	122
Db	9930	GCCCAAGCCCTTGAGATCAITTAACATGACATTTGAGCCTCCCATACATGTCATTCACA	9989
QY	123	GACTCACTTGGTTTCATTTGGGAAAGTTGGAAGATGATGAATATGATTTTGGAT	182
Db	9990	CTCTCATGAGGCAACATGGAGAAATGTCTCATATTGATCTTGAGGAAATGAAGCTTGGCA	10049
QY	183	TCCTTCACTCGGTTGGTTCATGATGAGTCGACCTCAAGAGTCCTCCAAATATCAAAGCC	242
Db	10050	TCCTTGATCTCATTTGTACCCCTCATGATCTCGGCCAACTTTTCCAAATCTCATATGG	10109
QY	243	GAATCACAAAATGAAATGTGATTTGAATTCATTTTGTCTAATGCAACAAACAGGGCATTC	302
Db	10110	GTCTCAAGGTTGCTCACTCTATCGAATCTCTCTTGGCTCAAGAGTTGAACAGGCATTC	10169
QY	303	ATAGCCTTTGTGTTTTAAAGCAAAAACATCTTCTCCGATTCATCCCATTCGCTCATCGGA	362
Db	10170	ATGGC-----TTGGGCATTGAGTTGGAGGTTGCGGTGATCAATCTCCGTCAAGGT	10230
QY	363	AGAGAAAAATTTTGAATTCATTTTCGCAATAGACCAAGCTCGAAATCGAATG	422
Db	10221	GTCCCGGTGATGCAAGCCATACATCCCAATCTCCAATATGGAAGCTCATAGCTTTG	10280
QY	423	AGGAAGATCCTCATATGAGTTTTTCCAATACATGTAATTCGACTCATTAACATAGTGGA	482
Db	10281	AGGTGAGTACACATTTTAATTTTCCAAGTGGAGTAGTTTGTGCCATTAACATGGAGCC	10340
QY	483	TGTGTAAATGAAATGACCCCTCATGCSCTAT	511
Db	10341	TTCCCTACATGGTTTCACCTCGTTTCGACAT	10369

Search completed: September 18, 2004, 18:55:07  
Job time : 971.103 secs

```

RESULT 15
US-10-437-963-9586/c
; Sequence 9586, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221) B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 9586
; LENGTH: 4002
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_1598C.1
US-10-437-963-9586

```

Query Match 3.3%; Score 70.2; DB 17; Length 4002;  
Best Local Similarity 48.9%; Pred. No. 1.6e-07;

	Matches	249;	Conservative	1;	Mismatches	249;	Indels	10;	Gaps	2;
Qy	4	TCCCAA	CTTTTAGGAATGGATCTTAAATTTT	TAGTTATAAGTTCAAAAGTTAGAAAAATCT	63					
Db	500	TCATAC	TTCCCGTAGGGCTCTAAGCATCTTTGGCAACCTCAAGATCGGTGTAGTTT	441						
Qy	64	TTACCA	AGAGCTTTAGTGCATTTGATGACATCCGTGAAAC-GGTGTACATGTCTCCGATG	122						
Db	440	GCCCCA	AGCGCTTGAGATCATTCACAATGACATTTGAGCCTCCCATACATGTCAATCACA	381						
Qy	123	GACTCA	CTTGGTTTCATTTCGAAAAGTTGCAAGAGTGATAGAAATATTGATTTGGAT	182						
Db	380	CTCTCA	TGAGGCAACATGGGAATGTCTCATTTGGATCTTTGAGGAAATGAAGCTTGGCA	321						
Qy	183	TCCTTCA	CTCGGTGGTCCCTTCATGAGTGACCTCAAGAGTCTCTCAAAATATCAAAGCC	242						
Db	320	TCCTTGT	ACTTGTACCTCATGGATCTCGGCCAATTTGTTCCAAATCTCATATGCG	261						
Qy	243	GAATCA	CAAAATGAAATGTGATTGAAATTCATTTTGTCTAATGCACAAAACAGGGCATTC	302						
Db	260	GTCTCA	AGGTGCTCACTATCGAACTCCTCTTGGCTCAAAGAGTTGAACAGGCGATTC	201						
Qy	303	ATAGC	CTTTGTGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCATTCGGCTCATCGGA	362						
Db	200	ATGSC	-----TTGGCATTGAGTTGGAGGTTTGGTGATCAATCTCCGTCAAAGGT	150						
Qy	363	AGAGAAA	TTTTTTCAAATCCATTTTCGACATATAGACCAAGCTCGAAATCCATGGAAATG	422						
Db	149	GTGCCG	GTATGGCAAGCCTACATCCCAATACTCORAATATGGAGGTCATAGCTTTG	90						
Qy	423	AGGAAGA	TCCTCATATGAGTTTTCCAATACATCTAATTCGACTCATTTAAACATAGTGGA	482						
Db	89	AGGTG	AGTAGACATTTTAAITTTTCCAAGTGGAGTAGTTTGTGCCATTAACATGGAGCC	30						
Qy	483	TGTGT	AATGAAATGACCCCTCATGCSCTAT	511						
Db	29	TTCCCT	ACATGGTTTCACCTCGTTCGCAT	1						

Search completed: September 18, 2004, 18:55:07  
Job time : 971.103 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 17:55:10 ; Search time 4963.9 Seconds  
(without alignments)  
12970.230 Million cell updates/sec

Title: US-09-892-635A-44

Perfect score: 2156

Sequence: 1 ggatcccaacttttaggaat.....agcctctcattcactatgg 2156

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_estl.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estcom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_pbg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gssl.\*  
29: gb\_gss2.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121.2	5.6	500	14	CD474210
2	96.6	4.5	790	28	BH021865
3	93.4	4.3	977	29	CG823740
4	93	4.3	964	28	CC412052

C	5	92.2	4.3	776	29	CG822306
C	6	92	4.3	832	29	CG818965
C	7	91.4	4.2	821	29	CG817604
C	8	87.6	4.1	787	29	CC804388
C	9	87.4	4.1	608	28	BZ778766
C	10	87.2	4.0	767	29	CG795738
C	11	87	4.0	873	29	CG8227901
C	12	86.6	4.0	707	28	BZ527252
C	13	85.6	4.0	970	29	CG878233
C	14	85.4	4.0	649	29	CG802722
C	15	85.4	4.0	660	29	CC741962
C	16	85.4	4.0	787	29	CG678004
C	17	85.4	4.0	800	28	BZ736364
C	18	85.4	4.0	820	29	CG796329
C	19	85.4	4.0	826	29	CG677989
C	20	85.4	4.0	856	28	BZ642762
C	21	85.4	4.0	989	29	CG365744
C	22	85.2	4.0	719	28	CC331635
C	23	84.6	3.9	821	29	CG701719
C	24	84	3.9	792	29	CC900423
C	25	83.8	3.9	681	28	BZ779455
C	26	83.8	3.9	709	28	BZ779456
C	27	83.8	3.9	822	29	CG739484
C	28	83.8	3.9	862	29	CG382637
C	29	83	3.8	732	28	BH794864
C	30	83	3.8	791	29	CG824360
C	31	82.2	3.8	538	28	BH780677
C	32	81.8	3.8	659	29	CG403457
C	33	81.8	3.8	890	29	CG817563
C	34	81.8	3.8	927	29	CG820967
C	35	81.4	3.8	733	29	CG815315
C	36	81.4	3.8	872	29	CG817875
C	37	81.2	3.8	295	14	CB340076
C	38	81.2	3.8	751	29	CG796506
C	39	81.2	3.8	795	29	CG816026
C	40	81.2	3.8	914	29	CG822083
C	41	80.6	3.7	656	29	CG995501
C	42	80.6	3.7	820	29	CG794988
C	43	80.4	3.7	867	29	CC994003
C	44	80	3.7	689	29	CC834718
C	45	79.8	3.7	396	29	AG236150

## ALIGNMENTS

RESULT 1	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
LOCUS	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
DEFINITION	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
ACCESSION	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
VERSION	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
KEYWORDS	EST				
SOURCE	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
ORGANISM	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
REFERENCE	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
AUTHORS	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
TITLE	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
JOURNAL	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003
COMMENT	CD474210	500 bp	mRNA	linear	EST 04-JUN-2003

CD474210 500 bp mRNA linear EST 04-JUN-2003  
nad03-10msl-909 Nad03 Nuphar advena cDNA clone nad03-10msl-909 5',  
mRNA sequence.  
CD474210 GI:31395478  
Nuphar advena  
Nuphar advena  
Nuphar advena  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; stem Magnoliophyta; Nymphaeaceae;  
Nuphar.  
1 (bases 1 to 500)  
dePamphilis, C., Carlson, J., Ma, H., Frohlich, M., Tanksley, S.,  
Leebens-Mack, J., Field, D., Arrington, J., Zahn, L., Kong, H.,  
Druckemiller, M., Landherr, L., Hu, Y., Ilut, D., Wall, K.,  
Plock, S., Chiorean, S., Albert, V., Doyle, J., Miller, W.,  
Oppenheimer, D., Soltis, D., Soltis, P. and Theissen, G.  
Generation of ESTs from early flower buds of Nuphar advena  
Unpublished (2002)  
Contact: Claude dePamphilis or James Leebens-Mack  
Mueller Laboratory  
Penn State University  
208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn  
State University, University Park, PA 16802, USA  
Tel: 814 863 6413

Fax: 814 865 9131  
Email: cwd3@psu.edu or jhl10@psu.edu  
The sequence provided is trimmed of vector and low quality regions.  
Full sequence and original trace file are available from the Plant  
Genome Network website (<http://pgn.cornell.edu>)  
Plate: nad03-10ms1 row: g column: 09  
Seq primer: M13F.  
Location/Qualifiers  
1. 500  
/organism="Nuphar advena"  
/mol\_type="mRNA"  
/db\_xref="taxon:77108"  
/clone="nad03-10ms1-909"  
/tissue\_type="flower buds"  
/dev\_stage="<= 2.5mm"  
/lab\_host="SOLR"  
/clone\_lib="Nad03"  
/note="Vector: pBluescript SK (+/-); Site 1: EcoRI;  
Site 2: XhoI; Only floral buds with diameter of 2.5 mm of  
less were used for RNA isolation. This is a directionally  
cloned, non-normalized library. Avg. insert length: 1134;  
Primers: M13F and M13R; Antibiotic: 50 ug/ml Ampicillin;  
Primary Titer: 286 pfu total; Amplified Titer: 3.2E10  
pfu/ml; Mass Excised Titer: 5E10 total; This library has  
been generated by the Floral Genome Project (FGP). The  
Floral Genome Project is funded by NSF's Plant Genome  
Research Program (DBI-0115684). More information about the  
project can be obtained at <http://fgp.bio.psu.edu>"

## ORIGIN

Query Match 5.6%; Score 121.2; DB 14; Length 500;  
Best Local Similarity 58.2%; Pred No. 4.1e-20;  
Matches 213; Conservative 0; Mismatches 153; Indels 0; Gaps 0;  
QY 119 GATGACTCACTGGTTTCATTCGGAAGAAGTTCGAAGAGTGCATPAGAATATTGATTT 178  
DB 2 GATAAATCGTTTTCCTTCATTTTAAATAATTCAAATTTATGAATGAAGAAGCAATCTT 61  
QY 179 GGATCTTTACTGGTGGTGCCTTATGAGTGACCTCAAGATCCTCCAAATATCAA 238  
DB 62 ATATTCTTTACCTTTGATGTACCTTCATGATGATGACGAGAAGATGATGCAAAATTCAT 121  
QY 239 AGCCGAATCAGAAATGGAATGGAATTCATTTTGTCTAATGCACAAACAGGGC 298  
DB 122 TCGTGTAGAACATTAAGAGGTGATTAATCTACCTTGACGGCACAATATAGCC 181  
QY 299 ATTCATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCGATTCATCCCATTCGCTCAT 358  
DB 182 ATTCATAGCTTTGGCGTTGAGTGAACAGTCTTCTTTCATCGGTCCATTCGCTCAA 241  
QY 359 CGAAGAGAAATTTTGAATCCATTTTCACATAGACCAAGCTCGAATCCATGGA 418  
DB 242 ATCAGTTGATGGTGGTGGTACTTTTCTGAAACAACTCTCCAAATTTTGAGATCTATAGA 301  
QY 419 AATCAGGAAGATCCATCATGATGTTTCCCAATACATGTAATTCGACTCATTAACATAGG 478  
DB 302 TTGTAAGATATTGTCACTCTACTTCCATAGCCATTAATTTGTTCGTCAAGAGAGG 361  
QY 479 TGGATG 484  
DB 362 AGGATG 367

## RESULT 2

BH021865 790 bp DNA linear GSS 29-JUN-2001  
LOCUS GH\_MB0001K07f Gossypium hirsutum L. Gossypium hirsutum genomic  
DEFINITION clone GH\_MB0001K07f, genomic survey sequence.  
ACCESSION BH021865  
VERSION BH021865.1 GI:14576153  
KEYWORDS GSS.  
SOURCE Gossypium hirsutum (upland cotton)  
ORGANISM Gossypium hirsutum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.  
1 (bases 1 to 790)  
Tomkins,J.P., Peterson,D.G., Yang,T.J., Main,D., Wilkins,T.A.,  
Paterson,A.H. and Wing,R.A.  
Development of Genomic Resources for Cotton (Gossypium hirsutum  
L.): BAC Library Construction, Preliminary STC Analysis, and  
Identification of Clones Associated With Fiber Development  
Unpublished (2001)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: [wing@clemson.edu](mailto:wing@clemson.edu)  
Seq primer: TAATPAGCACTCACTATAGGG  
Class: BAC ends  
High quality sequence stop: 532.  
Location/Qualifiers

JOURNAL  
COMMENT

1. 790  
/organism="Gossypium hirsutum"  
/mol\_type="genomic DNA"  
/cultivar="Maxxa"  
/db\_xref="taxon:3635"  
/clone="GH\_MB0001K07f"  
/tissue\_type="young leaves"  
/lab\_host="E. coli"  
/clone\_lib="Gossypium hirsutum L."  
/note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;  
For more details on library preparation, ordering clones  
and sequence analysis see  
[http://www.genome.clemson.edu/projects/stc/cotton/GH\\_MBb](http://www.genome.clemson.edu/projects/stc/cotton/GH_MBb)"

FEATURES  
source

Query Match 4.5%; Score 96.6; DB 28; Length 790;  
Best Local Similarity 54.4%; Pred. No. 9.4e-14;  
Matches 215; Conservative 0; Mismatches 179; Indels 1; Gaps 1;  
QY 4 TCCCAACTTTTGAAGATGGATCTTTAAATTTTATGTTTAAAGTTCAAGTTAGAAAATCT 63  
DB 281 TCCCATGATTTAGGCAAGCTTCAAAACATCTTCTTACCACTATTTCATTTGAATATGTT 340  
QY 64 TTACAAGAGCTTTGAGTCCATTGATGACATCCCGTGAACGG-TGTACATCTCTCCGATG 122  
DB 341 TTCCTATAATATCTTAATCCATTGATGATAATAGTAAACCGATGGACATCGCCTTGATA 400  
QY 123 GACTCACTTGGTTTTCATTCGGAAAAGTTTCGAAAGAGTGATAGAATAATTGATTTTGGAT 182  
DB 401 TCTTCTCAGGCTTCATTTTGAAGTTTCATAATTTAAAGTACGAATGCCAACCTTAGAA 460  
QY 183 TCTTTCTACGTGGTGGTCCCTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAGCC 242  
DB 461 TCTTTTACTTGGCTGGTACCTTCATGTGAGCTTCAAGCTTATCACATATCTCTCTTGCA 520  
QY 243 GAATCAGAAATGAAATGATGAAATTCATTTTGTCTAATGACAAAAAGGSCATTC 302  
DB 521 TTGAATCAAGATGAACCTCTACTATATATGCTGCTAGGCCAAGTACATATAAGCGTGC 580  
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCATCGGA 362  
DB 581 ATAGCTTGGCATTTAACTAAACACTTGTCTATCTTCTCTCATTCACATATATTGCTC 640  
QY 363 AGAGAAATTTTGAATCCATTTTCGACATAGA 397  
DB 641 TTTGGAATAATGAGAACT 675

## RESULT 3

CG823740/c  
LOCUS CG823740 977 bp DNA linear GSS 18-NOV-2003  
DEFINITION SOYDB38TH LargeInsertSoybeanGenLib Glycine max genomic clone



QY 4 TCCCAACTTTTGGAGATGATCTTAAATTTTAGTTTAAAGTTAGAAAATCT 63  
 |||  
 Db TCATATGCTTTGGAGAGATCTAAGGAATGTGAGTGAATCCACATCGATACATTA 150  
 |||  
 QY 64 TTACAAGAGCTTTGAGTCCATGATGACATCCGTAACCGG-TGTACATGCTCCGATG 122  
 |||  
 Db 151 AAACCAAGTCCCTTGAGCTCATCTACATCTCATTTAAACCGGTAAACATGTCGACACA 210  
 |||  
 QY 123 GACTCACTTGTTCATTCGAAAGATTCGAAAGATGATGATGATGATGATGATGAT 182  
 |||  
 Db 211 CTTTCATCCTCTTCATGAGAAATGCTCAAACTTTCCTTTGCGTACATACAACTAGCA 270  
 |||  
 QY 183 TCTTTCACTCGGTTGGTCCCTTCATGAGTACCTCAAGAGTCCCTCAAAATATCAAAAGCC 242  
 |||  
 Db 271 CTCCTTCAAAATGATGTTCTTCATGATGATGATGATGATGATGATGATGATGATG 330  
 |||  
 QY 243 GAATCACAAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 302  
 |||  
 Db 331 GTTGTGAGATTTTGTGATGCGGTGAACCTCAATTTATGTTCTAAAGCATCATAGATACATTC 390  
 |||  
 QY 303 ATAGCCTTGTGTTTAAAGCAAAACATCTCTCCGATTCATCCCATCTCCCTCATCGGA 362  
 |||  
 Db 391 ATAG-CTTGATCATGTGATGAGAAATTTTCATATCAATTTGGTGAATGTTCACTCTCTTT 449  
 |||  
 QY 363 AGAGAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGAAATCCATGGAATG 422  
 |||  
 Db 450 TAAATACAA-----CAATCCATCTTTACATATAGGCCAAATCTTTTCCCATAGCTTTA 504  
 |||  
 QY 423 AGAAGATCCCTCATATGATGTTTCCATATACATGATGATGATGATGATGATGATGATG 482  
 |||  
 Db 505 AGATGATCATGCTGCTTTATTTTCCATAGTGTGATGATGATGATGATGATGATGATGATG 564  
 |||  
 QY 483 TGTGTATG 491  
 |||  
 Db 565 TTCCCAATG 573

RESULT 5  
 CG822306/C  
 LOCUS  
 DEFINITION SOYEA21TV LargeInsertSoybeanGenLib Glycine max genomic clone  
 H45P23:MTPI5C17, genomic survey sequence.  
 CG822306  
 VERSION CG822306.1 GI:38281855  
 KEYWORDS GSS  
 SOURCE Glycine max (soybean)  
 ORGANISM Glycine max

REFERENCE  
 AUTHORS Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,  
 Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.  
 TITLE End sequencing of BACs comprising a provisional minimal tiling path  
 from a fingerprint physical map of soybean (Glycine max) cultivar  
 Forrest

JOURNAL  
 COMMENT Unpublished (2003)  
 Other GSSs: SOYEA21TH  
 Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
 The Center of Excellence in Soybean Research, Teaching and  
 Outreach, Southern Illinois University at Carbondale and Plant  
 Genomics, The Institute for Genomic Research  
 Room 176, Ag Building, Mail Code 4415, Carbondale, IL 62901-4415,  
 USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 618 453 1797  
 Email: ga4082@siu.edu, jleshultz@siu.edu, cdtown@igrr.org (URL:  
 http://bioinformatics.siu.edu)  
 Clones approximating a minimum tiling path were re-arrayed from the  
 library master plates prior to sequencing.  
 For purposes of clone identification each clone name is a  
 concatenation of the original clone location and its new location  
 in the re-arrayed sequencing plates.

Seq primer: GTAATACGACTCACTATATAGGC  
 Class: BAC ends  
 FEATURES  
 source  
 1..776  
 Location/Qualifiers  
 /organism="Glycine max"  
 /mol\_type="genomic DNA"  
 /cultivar="Forrest"  
 /db\_xref="taxon:3847"  
 /clone="H45P23:MTPI5C17"  
 /clone\_lib="LargeInsertSoybeanGenLib"  
 /note="Organ: Leaves; Vector: pCD04541 (pBELOBAC11 Ecori  
 clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
 cv. Forrest seeds were grown in greenhouse for fourteen  
 days. Nuclei were isolated and embedded in agarose.  
 restriction digested with Hind III BamHI or EcoRI, large  
 size DNA fragments were ligated in vector v41 (pCD04541)  
 and electro transformed in DH10a cells. About 90,000  
 clones from BAC libraries were fingerprinted with HindIII  
 and Hae III. Version 2 (automatic build) Contigs were  
 built from 78,001 fingerprints. Contigs were manually  
 examined to find the best non redundant tile path through  
 the contigs representing about 13,000 clones. The clones  
 were end sequenced."

## ORIGIN

Query Match 4.3%; Score 92.2; DB 29; Length 776;  
 Best Local Similarity 51.8%; Pred. No. 1.3e-12;  
 Matches 277; Conservative 0; Mismatches 233; Indels 25; Gaps 2;  
 QY 1 GGATCCCAACCTTTAGGAATGGATCTTAAATTTTAGTTATATAGTTCAAAAGTTAGAAAA 60  
 Db 568 GGTGGCCATCTCTACTTAAACATCTTAAACATCTTATTTATGAGATCTCTGTTTGA 509  
 |||  
 QY 61 TCCTTACCAGAGCTTTGAGTCCATTCATGATGATCCGTAAC--GGTGATCATGCTCCG 119  
 |||  
 Db 508 GTTCTTCTTAATGATGCAAGATGATTAACATATATGTGTGAATCTTTCTGTGATATCTTGT 449  
 |||  
 QY 120 ATGACTCACTTGTGTTTCATTCGGAAAGTTTCGAAAGAGTGCATAAGAAATATGATTTTG 179  
 |||  
 Db 448 ATACTTCATTTGTCTTCATCCCTAAATATTCATCTACTCATGATTAAGATTTATCTCTA 389  
 |||  
 QY 180 GATCTTTTCACTCGGTTGGTGGCTTCATGATGATGATGATGATGATGATGATGATGATG 239  
 |||  
 Db 388 GATCTTTTGACATCAGTTGTTCCCTCATGTGTAATCTGTAGAGTGTCCACATATCTCTTA 329  
 |||  
 QY 240 GCCGAATCACAAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 299  
 |||  
 Db 328 GCATCTTACAAATTTGAAACCCCTAAATATATTCATCCATTCCTAGGGCAGAAATATAATA 269  
 |||  
 QY 300 TTCATAGCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTCGCTCATC 359  
 |||  
 Db 268 TTTTATGCTTTAAATATATCTGATGATGATGATGATGATGATGATGATGATGATGATG 209  
 |||  
 QY 360 GG-----AAGAGAAAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 395  
 |||  
 Db 208 GGTTCCTCTATTGTATTTCAGCCACCATGTTGGGTACATAAGGTCCCACTTCTATGGCA 149  
 |||  
 QY 396 GACCAAGCTCGAAATCCATGGAATGAGAGAGATCTCTCATATGATGATGATGATGATGATG 455  
 |||  
 Db 148 TCCCAAAATGTTTAAAGTCTATTGCTCCTCAATGAAATTTGCAATCGGGTTTCCAGTAGTGG 89  
 |||  
 QY 456 TAAATTCGACTCATTTAAACATAGTGGATGTTGTAATGAAATGACCTCATGCTCA 510  
 |||  
 Db 88 TAACTCTCCATTAATAAATAGGAGGCTATTATTGATTTCTCTCTGGAATA 34  
 |||

RESULT 6  
 CG818965  
 LOCUS  
 DEFINITION SOYED06TV LargeInsertSoybeanGenLib Glycine max genomic clone  
 H5G03:MTPI5B12, genomic survey sequence.  
 CG818965  
 ACCESSION CG818965  
 VERSION CG818965.1 GI:38275242

```

KEYWORDS
SOURCE      Glycine max (soybean)
ORGANISM    Glycine max

REFERENCE   1 (bases 1 to 832)
AUTHORS    Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,
            Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.
TITLE      End sequencing of BACs comprising a provisional minimal tiling path
            from a fingerprint physical map of soybean (Glycine max) cultivar
            Forrest
JOURNAL     Unpublished (2003)
COMMENT     Other_GSSs: SOYED06TH
            Contact: Chris Town, J. L. Shultz and D. A. Lightfoot
            The Center of Excellence in Soybean Research, Teaching and
            Outreach, Southern Illinois University at Carbondale and Plant
            Genomics, The Institute for Genomic Research
            Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,
            USA and 9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 618 453 1797
            Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL:
            http://bioinformatics.siu.edu)
            Clones approximating a minimum tiling path were re-arrayed from the
            library master plates prior to sequencing.
            For purposes of clone identification each clone name is a
            concatenation of the original clone location and its new location
            in the re-arrayed sequencing plates.
            Seq primer: GTAATACGACTCACTATAGGGC
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..832
                     /organism="Glycine max"
                     /mol_type="genomic DNA"
                     /cultivar="Forrest"
                     /db_xref="taxon:3847"
                     /clone="H50G01-MTP15B12"
                     /clone_lib="LargeInsertSoybeanGenLib"
                     /notes="Organ: Leaves; Vector: pCUD04541 (pBELOBAC11 EcoRI
                     clones); Site:1: BatXI; Soybean (Glycine max (L.) Merr.)
                     cv. Forrest seeds were grown in greenhouse for fourteen
                     days. Nuclei were isolated and embedded in agarose,
                     restriction digested with Hind III BamHI or EcoRI, large
                     size DNA fragments were ligated in vector V41 (pCUD04541)
                     and electro transformed in DH10a cells. About 90,000
                     clones from BAC libraries were fingerprinted with HindIII
                     and Hae III. Version 2 (automatic build) Contigs were
                     built from 78,001 fingerprints. Contigs were manually
                     examined to find the best non redundant tile path through
                     the contigs representing about 13,000 clones. The clones
                     were end sequenced."

ORIGIN
Query Match      4.3%; Score 92; DB 29; Length 832;
Best Local Similarity 56.1%; Pred. No. 1.4e-12;
Matches 193; Conservative 0; Mismatches 150; Indels 1; Gaps 1;

QY 19 ATGGATCTTAAATTTAGTATTAAAGTTCAAGTTAGAAAAATCTTTACCAAGAGCTTTG 78
Db 171 AAGCATCTTAAACATTTATTGACAAGATCTTCATTTGGAAATATTTTCTTAAAGATGCA 230
QY 79 AGTCATTGATGACATCCGTTGAAC--GGTGTACATGTCCTCCGATGACTCACTGGTTTC 137
Db 231 AGATGATTTTATTATGTTGGGTAATCTCTTTTGGCATGTCCTTTGTATAGTTTCATTAGATTC 290
QY 138 ATTTCGGAAAAGTTTCGAAGAGTGTCATAAGAAATATTTGATTTTGGATCTTTCACTCGGTG 197
Db 291 ATTCTAAATAACTCATATTCATGAGTTAAAGTATTTATCCTAGATCTTTTATCATCTGTT 350
QY 198 GTCCCTTCATGAGTGACCTCAAGAGTCTCCTCAAAATATCAAAAGCCGAATCACAAATGAA 257
Db 351 GTCCCTTCATGAGTTACTTGTAGGTTATCCACATATCTTTTGCATGTTTGCATTTTGAC 410

```

```

QY 258 ATGTGATTGAATTCATTTTGTCTTAATGCACAAACAGGSCATTCATAGCCTTTGTGTTT 317
Db 411 ACCCTAAAGATTATCATCTATTCCTAATAGTAGATGTAATGAATATTTTGGCTTTAAATG 470
QY 318 AAAGCAAAACATCTTCTCCGATTCATCCATCGCTCATCGG 361
Db 471 TATTGAACTAATCTCTTTCTCTCTCTCACTCCATTAATCCATAGG 514

```

```

RESULT 7
CG817604/c
LOCUS      CG817604
DEFINITION H32A21:MTP13A5, genomic survey sequence.
ACCESSION  CG817604
VERSION    CG817604.1 GI:38272539
KEYWORDS   Glycine max (soybean)
SOURCE     Glycine max
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.

```

```

REFERENCE   1 (bases 1 to 821)
AUTHORS    Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,
            Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.
TITLE      End sequencing of BACs comprising a provisional minimal tiling path
            from a fingerprint physical map of soybean (Glycine max) cultivar
            Forrest
JOURNAL     Unpublished (2003)
COMMENT     Other_GSSs: SOYDQ03TH
            Contact: Chris Town, J. L. Shultz and D. A. Lightfoot
            The Center of Excellence in Soybean Research, Teaching and
            Outreach, Southern Illinois University at Carbondale and Plant
            Genomics, The Institute for Genomic Research
            Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,
            USA and 9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 618 453 1797
            Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL:
            http://bioinformatics.siu.edu)
            Clones approximating a minimum tiling path were re-arrayed from the
            library master plates prior to sequencing.
            For purposes of clone identification each clone name is a
            concatenation of the original clone location and its new location
            in the re-arrayed sequencing plates.
            Seq primer: GTAATACGACTCACTATAGGGC
            Class: BAC ends.

```

```

FEATURES             Location/Qualifiers
     source           1..821
                     /organism="Glycine max"
                     /mol_type="genomic DNA"
                     /cultivar="Forrest"
                     /db_xref="taxon:3847"
                     /clone="H32A21:MTP13A5"
                     /clone_lib="LargeInsertSoybeanGenLib"
                     /notes="Organ: Leaves; Vector: pCUD04541 (pBELOBAC11 EcoRI
                     clones); Site:1: BstXI; Soybean (Glycine max (L.) Merr.)
                     cv. Forrest seeds were grown in greenhouse for fourteen
                     days. Nuclei were isolated and embedded in agarose,
                     restriction digested with Hind III BamHI or EcoRI, large
                     size DNA fragments were ligated in vector V41 (pCUD04541)
                     and electro transformed in DH10a cells. About 90,000
                     clones from BAC libraries were fingerprinted with HindIII
                     and Hae III. Version 2 (automatic build) Contigs were
                     built from 78,001 fingerprints. Contigs were manually
                     examined to find the best non redundant tile path through
                     the contigs representing about 13,000 clones. The clones
                     were end sequenced."

```

```

ORIGIN
Query Match      4.2%; Score 91.4; DB 29; Length 821;
Best Local Similarity 55.5%; Pred. No. 2.1e-12;

```

Matches 196; Conservative 0; Mismatches 156; Indels 1; Gaps 1;

QY 1 GGATCCCAACTTTAGGAATGATCTTAAATTTTAGTTATAAGTTCAAGTTAGAAAA 60  
|||  
Db 571 GGTGCGCAATCTCTATTAAAGCACTTTAAAGACTTTATTATTAGATCTTCAATTTGAAAA 512  
|||  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTGATGATCATCCGTTGAACCGTGTA-CATGTCCTCG 119  
|||  
Db 511 GTTTTCTAGAGATGCAAGATGATTAACTATGTGTGGAACCTCTTTGCAATGCTTGT 452  
|||  
QY 120 ATGACTCACTGTTTCATTCGGAAGTTGGAAGAGTGCAFAAGAAATATGATTTTG 179  
|||  
Db 451 ATACTTTTATTACATTCATCTAAAAGATTCAATTCACGAGTTAAAGTCTTTATCCTA 392  
|||  
QY 180 GATTCCTTCACTCGTGGTGCCTTCATGAGTGACCTCAAGAGTCCCTCCAAATATCAAAA 239  
|||  
Db 391 GATCTTTTAACTCTGTTGTGCCCTTCAGTGTACTTGTATGTTATCCCAATATCCCTTA 332  
|||  
QY 240 GCCGAATCACAAATTTGAATGTGATTTGAATTCATTTTGTCTAATGCACAAAACAGGGCA 299  
|||  
Db 331 GCACCTTTTACAATTTGAAACCTTAAAGTATTCACTATTCTTAAGGCAGATGTAATAA 272  
|||  
QY 300 TTCAATAGCTTTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATC 352  
|||  
Db 271 TTTTGGCCTTTAAATATATTGTAATCTTCTTCTTCTTCTCCTCAGTCCCAATC 219  
|||

RESULT 8  
CC804388  
LOCUS  
DEFINITION  
ih28all.g1 WGS-ZmaysU (DH10b or JM107MA2) Zea mays genomic clone  
ih28all, genomic survey sequence.

ACCESSION  
CC804388  
VERSION  
CC804388.1 GI:32629811  
KEYWORDS  
GSS.

SOURCE  
Zea mays  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 787)  
Rabinowicz,P.D., O'Shaughnessy,A.L., Palmer,L., Balija,V.,  
Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S.,  
Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.  
Genomic shotgun sequences from Zea mays (unfiltered)  
Unpublished (2003)  
Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: ih28 row: a column: 11  
Seq primer: -21M13UnivRev  
Class: shotgun  
High quality sequence stop: 787.

FEATURES  
source  
1..787  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ih28all"  
/lab\_host="DH10b or JM107MA2"  
/clone\_lib="WGS-ZmaysU (DH10b or JM107MA2)"  
/note="Organ: immature ears; Site\_1: Xba I; Site\_2: Xba I;  
The vector was digested with XbaI and one nucleotide was  
added by fill in in the recessive 3' end. The genomic DNA  
was nebulized, end repaired, adaptor ligated and size  
fractionated using sephadex. The resulting fragments were  
between 0.8 and 3 kb and were cloned into the vector (.x/y  
reads in M13mp19, .b/g reads in pUC19). The ligation was

## ORIGIN

transformed into DH10b or JM107MA2."

Query Match 4.1%; Score 87.6; DB 29; Length 787;  
Best Local Similarity 51.6%; Pred. No. 2e-11;  
Matches 252; Conservative 0; Mismatches 229; Indels 7; Gaps 2;

QY 4 TCCCAACTTTTAGGAATGATCTTAAATTTTAGTTATAAGTTCAAGTTAGAAAAATCT 63  
|||  
Db 260 TCATATTTCTCGGAAGTGATCTAAGGAACCTTGTGTGGAATCCACATCCGGTACATTA 319  
|||  
QY 64 TTACCAAGAGCTTTGAGTCCATTGATGACATCCGTGAAACG-GTGTACATCTCTCCGATG 122  
|||  
Db 320 AACCAAGTCTTTGAGTTCATTACTATCTCTCAATCAACCGATTGAACATATCGATACA 379  
|||  
QY 123 GACTCACTCTGGTTTCATTCGGAAGTTGGAAGAGTGATGAAGAAATATTCATTTTGGAT 182  
|||  
Db 380 CTCTCATCTCTTTTCAATAATGCTCAAACTTCCCTTTGCACACATATAGTTTGGCA 439  
|||  
QY 183 TCTTCACTCGTTGTGGTGCCTTCATGAGTGACCTCAAGAGTCTCCCAATATCAAAAGCC 242  
|||  
Db 440 CTCCTTCAAAATTTGAGTTCCTTCATGAATTTCCATTAGTTTATCCAAATCTCATGAGCG 499  
|||  
QY 243 GAATFCACAAAATTTGAATGTGATTCGAATTCATTTTGTCTAATGCACAAAAACAGGCGCATTC 302  
|||  
Db 500 GTTGTGAGATTTTGTATACGATTGAACCTCAATTTATATCAAGAGCATCATAAAGACATTC 559  
|||  
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATTCGCTCATCGGA 362  
|||  
Db 560 ATGCTTGTGATCAATTTGTGAGGATATTCCTATATCACTAAGCGATGGTTTCATCTTCC--- 616  
|||  
QY 363 AGAGAAAATTTTGAATTCATTTTCGACATAGACCAAGCTCGAAATCCATCGAATG 422  
|||  
Db 617 ---TTCATACACAAATCCATCCCTAACAAATTTGGCCAAATTTTCCACCCATTTGCTCTA 673  
|||  
QY 423 AGGAAGATCTCTCATATGAGTTTTCCAATACATATGTAATTCGACTCATTTAAACATAGGTGGA 482  
|||  
Db 674 AGATGATTCGACATCTTATTTTCCATAATCAATAATTTTCCATCAAGTGGGTGGC 733  
|||  
QY 483 TGTGTAAT 490  
|||  
Db 734 TTCCCAAT 741  
|||

RESULT 9  
BZ778766/c  
LOCUS  
DEFINITION  
ih84dl0.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone  
ih84dl0, genomic survey sequence.

ACCESSION  
BZ778766  
VERSION  
BZ778766.1 GI:28956209  
KEYWORDS  
GSS.

SOURCE  
Zea mays  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 608)  
Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,  
Katzenburger,F., King,L., Miller,B., Muller,S.,  
Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.  
Genomic shotgun sequences from Zea mays (methyl-filtered)  
Unpublished (2002)  
Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: ih84 row: d column: 10  
Seq primer: -21M13UnivFwd  
Class: shotgun

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

High quality sequence stop: 608.

# FEATURES

source  
1. .608  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="ih84d10"  
/lab\_host="DH5a"  
/clone\_lib="WGS-ZmaysF (DH5a methyl filtered)"  
/note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;  
The vector was digested with XbaI and one nucleotide was  
added by fill in the recessive 3' end. The genomic DNA  
was nebulized, end repaired, adaptor ligated and size  
fractionated using sephadex. The resulting fragments were  
between 0.8 and 3 kb and were cloned into the vector (x/y  
reads in M13mp19, b/g reads in pUC19). The same ligation  
was transformed into DH5a."

# ORIGIN

Query Match 4.1%; Score 87.4; DB 28; Length 608;  
Best Local Similarity 50.9%; Pred. No. 2.2e-11;  
Matches 259; Conservative 1; Mismatches 242; Indels 7; Gaps 2;  
QY 4 TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTTAAAGTTCAAGTTAGAAAATCT 63  
Db 512 TCATCTTATCCGAAGAGATCTAAGAACTTAGTGAATCCACATCTGGTACATTA 453  
QY 64 TTACCAAGAGCTTTGAGTCATTTGATGACATCCGCGTGAACGG-TGTACATGTCCTCGATG 122  
Db 452 AAACCAAGTCCCTTGAGTTCACTCACTCTCTCACTTTAACCGGTTGAACATATCAGATACA 393  
QY 123 GACTCACTGGTTTCATTCGGAAGAGTTCGAAGAGTGCATAAGATATGATTTGGAT 182  
Db 392 CTTTCATCTCTTTCATATGAATTCCTCAAAATTTTCTTGACACATATAGCTTGGCA 333  
QY 183 TCTTCACATCGGTTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCC 242  
Db 332 CTTTTCACAAATGTAGTTCCTTCATGAATTCATAGCTTAATCCAAATCICATGAGCG 273  
QY 243 GAATCAAAATGAATGTGATTAATTCATTTTCTTAATGCACAAACAGGGGCAATTC 302  
Db 272 GTTGTGAAGATCTTGTATGACGGTTGAATCACTTATATCAAGAGCATCATAGAATACATTC 213  
QY 303 ATAGCCCTTGTGTTTAAAGCAAAACATCTTCTCGAATTCATCCCATCGCTCATCGGA 362  
Db 212 ATGCTTGATCATTTGTAAGATGTTTTTCATTCGCAACAGATGGTTTCATCTTCTTTC 153  
QY 363 AGAGAAAATTTTGAATCCATTTTCGACAAATAGACCAAGCTCGAAATCCATGGAAATG 422  
Db 152 AACACAA-----CAAAACCATCTTGACAAATTTGGCCAAATTTTCCACCCATTTGCTTA 99  
QY 423 AGGAAGATCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAACATAGTGGGA 482  
Db 98 AGATGCATTAACATCTTATTTCCAAATAGTCGTAGTTGTGCCATCAAAATCCGGTGGC 39  
QY 483 TGTGTAATGAATGACCCCATGCSCTAT 511  
Db 38 TTCCCAATGTCACATTTCTCGAGGCGTCT 10

# RESULT 10

CG795738  
LOCUS  
DEFINITION  
ZMMBBb0329L23.f ZMMBBb Zea mays subsp. mays genomic clone  
ZMMBBb0329L23 5', genomic survey sequence.  
ACCESSION  
CG795738  
VERSION  
CG795738.1 GI:38212973  
KEYWORDS  
GSS.  
SOURCE  
Zea mays subsp. mays (maize)  
ORGANISM  
Zea mays subsp. mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

# REFERENCE

1. (bases 1 to 767)  
AUTHORS  
Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.,  
and Wing, R.  
TITLE  
Sequencing of the maize genome  
JOURNAL  
Unpublished (2003)  
COMMENT  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: T7  
BACKWARD: M13r  
Plate: 0329 row: L column: 23  
Seq primer: T7  
Class: BAC ends.

# FEATURES

Location/Qualifiers  
1. .767  
/organism="Zea mays subsp. mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/sub\_species="mays"  
/db\_xref="taxon:4578"  
/clone="ZMMBBb0329L23"  
/lab\_host="DH10B"  
/clone\_lib="ZMMBBb"  
/note="Vector: pBelobAC11; Site 1: HindIII; Site 2:  
HindIII; Zea mays L. ssp. mays"

# ORIGIN

Query Match 4.0%; Score 87.2; DB 29; Length 767;  
Best Local Similarity 52.9%; Pred. No. 2.5e-11;  
Matches 258; Conservative 0; Mismatches 223; Indels 7; Gaps 3;  
QY 7 CAACITTTAGGAATGGATCTTAAATTTTAGTTTAAAGTTCAAGTTAGAAAATCTTTTA 66  
Db 158 CATTTCTCTGGAAGGGATCTAAGGAACCTTATGAGTAAATCCACATCCGGTACATCAAGA 217  
QY 67 CCAAGAGCTTTGAGTCCATTGATGACATCCGCGAAACGG-TGTACATGTCCTCGATGGAC 125  
Db 218 CCAATCCCTTGAGCTCATTTACTATCTCATTTAACGGTTGACATGTCAGTACATTT 277  
QY 126 TCATTTGGTTTCATTCGGAAGAGTTGCAAGAGTGCATAAGAAATATTGATTTTGGATTC 185  
Db 278 TCAATTTCTTCATATAAATTTGCTCAATTTCCCTTCGCACACATATAATTTGACACTC 337  
QY 186 TTCACCTCGGTTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCGGA 245  
Db 338 TTCACAAATTTGATGTTCCCTTCATGAATTTCCATTTAGCTTCATCCAAATATCATAAAGCGGTT 397  
QY 246 TCACAAATTTGAATGATGATTGAATTCATTTTGTCTAATGCACAAACAGGGCATTCATA 305  
Db 398 GTGAGATCTTGTGATCTTATTTGAACTCAGTGTCTAGAGCATCATAGACATTCATG 457  
QY 306 GCGTTTGTGTTTAAAGCAAAACATTTCTTCTCGATTCATCCCATTCGCTCATFCGGAAGA 365  
Db 458 GCTTGACCATGATTAGAACATTTTTCATTATC---ACTAGTCAATAGTTTCATC--TTGC 511  
QY 366 GAAATTTTGAATTCATTTTCGACATAGACCAAGCTCGAAATCCATGGAATGAGG 425  
Db 512 TTCAATACAAACAAATTCATCTCTGGCAATTTGCCAAACCTTTCCACCCATTTGCTCTATGA 571  
QY 426 AAGATCCTCATATGAGTTTCCAAATACATGTAATTCGATTCATTAACATAGTGTGATGT 485  
Db 572 TGCATTTGACATCTTATTTTCCAAATAGTCAATATTTGTTCCCATCAAAATGTGATGCTTT 631  
QY 486 GTAATGAA 493  
Db 632 CCAATGCA 639

```

RESULT 11
LOCUS      CG227901                      873 bp    DNA        linear    GSS 22-AUG-2003
DEFINITION OGOBU24TH ZM 0.7.1.5_KB Zea mays genomic clone ZMMBMA0682D23,
            genomic survey sequence.
ACCESSION  CG227901
VERSION     CG227901.1  GI:34127787
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 873)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE     Consortium for Maize Genomics
JOURNAL   Unpublished (2002)
COMMENT   Other_GSSs: OGOBU24TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: IR
            Class: sheared ends.
FEATURES             source
            1..873
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBMA0682D23"
            /clone_lib="ZM 0.7.1.5_KB"
            /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"
ORIGIN
Query Match      4.0%; Score 87; DB 29; Length 873;
Best Local Similarity 51.5%; Pred. No. 2.8e-11;
Matches 252; Conservative 0; Mismatches 230; Indels 7; Gaps 2;

QY  4  TCCCACTTTAGGAATGGATCTTAAATTTTGTAGTTAAGTTCAGAGTTAGAAAATCT 63
DB  21 TCATATTTCTCGAAGTGTCTAAGGAACCTTGTAGTGAATCCACATCCGGTACATTA 80
QY  64 TTACCAAGAGCTTTGAGTCCATTCATGACATCCGTTGAACGG-TGTACATGTCCTCCGATG 122
DB  81 AACCACAGTCCCTTGAGTTCAATTCACATCTCATTTAACCGGTTGAACATATCGGATACA 140
QY  123 GACTCACTTGTTTCATTCGGAAGAAGTTCGAAAGAGTGCATAGAATATTGATTTTGGAT 182
DB  141 CTCTCATCTCTTTCATAATAAATGCTCAAACCTTCCCTTTGCACACATATAGTTTGCA 200
QY  183 TCCTTCACTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCCCTCCAAAATATCAAAAGCC 242
DB  201 CTCTTCACAAATTTAGTTCCTTCATGAATTTCCATTAGCTTAATPCCAAATCTCATGAGCA 260
QY  243 GAATCACAAAATGAAATGTGATTAATTTTGTCTAATGCACAAAACAGGGCATTC 302
DB  261 GTTGTGAGATCTTGATGCGATTTGAATCACTTATATATCAAGAGCATCATGAGACATTC 320
QY  303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCGGAATTCATCCCATTCGCTCATCGGA 362
DB  321 ATGCGTTGATCAATTTGTAAGGATATCTCATTTGTCACTAACCGGATGTTTCATCTCTTC 380
QY  363 AGAGAAATTTTGAATATCCATTTTCACAAATAGACCAAGCTCGAATCCATGGAATG 422
DB  381 AACACAA-----CAATTCATCCCTGACAAATGGCCAAATTTTTCACCCCATTCGCTTTA 434

```

```

QY  423 AGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAACATAGGTGGA 482
DB  435 AGATGCATTGACATCTTATTTTCTAATAGTCATAGTTATTTCTCATCAAAATCGGTGGC 494
QY  483 TGTGTAATG 491
DB  495 TTCCCAATG 503

RESULT 12
LOCUS      BZ527252                      707 bp    DNA        linear    GSS 16-DEC-2002
DEFINITION OGA160TC ZM2_0.7.1.5_KB Zea mays genomic clone ZMMBMA0069J24,
            genomic survey sequence.
ACCESSION  BZ527252
VERSION     BZ527252.1  GI:27067787
KEYWORDS   GSS.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 707)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE     Consortium for Maize Genomics
JOURNAL   Unpublished (2002)
COMMENT   Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             source
            1..707
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBMA0069J24"
            /clone_lib="ZM2_0.7.1.5_KB"
            /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"
ORIGIN
Query Match      4.0%; Score 86.6; DB 28; Length 707;
Best Local Similarity 52.8%; Pred. No. 3.5e-11;
Matches 259; Conservative 0; Mismatches 224; Indels 7; Gaps 3;

QY  4  TCCCACTTTAGGAATGGATCTTAAATTTTGTAGTTAAGTTCAGAGTTAGAAAATCT 63
DB  120 TCATACTTCTCTGGAAGAGATCTAAGGAACCTTATGAGTGAATCCACATCAGGTACATTA 179
QY  64 TTACCAAGAGCTTTGAGTCCATTCATGATGACATCCGTTGAACGG-TGTACATGTCCTCCGATG 122
DB  180 AAACCAAGTCCCTTGAGTTCAATTCATCTCATCTCATTTAAACCGGTTGAACATATCAGATACA 239
QY  123 GACTCACTTGTTTCATTCGGAAGAAGTTGGAAGAGTGCATAGAATATTGATTTTGGAT 182
DB  240 CTTTTCATCTTATTTCAATGAATTTGCTTAAATTTTCCTTTGCACACATATAGCTTTGCA 299
QY  183 TCCTTCACTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCCCTCCAAAATATCAAAAGCC 242
DB  300 CTCTTCATAATTTAGTTCCTTCATGAATTTCCATTAGCTTAATPCCAAATCTCATGAGCA 359
QY  243 GAATCACAAAATGAAATGTGATTAATTTTGTCTAATGCACAAAACAGGGCATTC 302
DB  360 GTTGTGAGATCTTGTGATACGAATTTGAATCTCATTTATGTCAAGAGCATCATGAGATACATTC 419

```



```

QY 303 ATAGCCTTTGTTTAAAGCAAAACATCTCTCCGATTCATCCATTCCTGCTCATCGGA 362
Db |||||
QY 420 ATGCTTGATCATTTGAGAAATGTTTATTTGTC---ACTAGCCGATGTTTATCTTC 475
Db |||||
QY 363 AGAGAAATTTTGAATCCATTTTCGCAATAGACCAAGCTGAAATCCATCGAAATG 422
Db |||||
QY 476 TTTCACACACAT--GAACCCATCTTCGACAATTCGCCAAATTTTCCACCCATTCCTCTA 533
QY 423 AGAAGATCCCTCATATGAGTTTTCCAATACATGTAATTCGACTCATTTAAACATAGTTGA 482
Db |||||
QY 534 AGATGCATTCGATCTTTATTTTCCAAATAGTAATGTTTATTTCCATCAAAATGCGGTGGC 593
QY 483 TGTGTAATG 491
Db |||||
QY 594 TTCCCAATG 602
Db |||||

RESULT 13
CG878233
LOCUS ZMMBB0496H16f ZMMBBb (HindIII) Zea mays subsp. mays genomic clone
DEFINITION ZMMBB0496H16 5', genomic survey sequence.
ACCESSION CG878233
VERSION CG878233.1 GI:38608836
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 970)
AUTHORS Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
TITLE Sequencing of the maize genome at PCR (2003c)
JOURNAL Unpublished (2003)
COMMENT Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
130 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: T7
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..970
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/lab_host="ZMMBB0496H16"
/clone="ZMMBBb (HindIII)"
/note="Vector: pCUG1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match 4.0%; Score 85.6; DB 29; Length 970;
Best Local Similarity 52.7%; Pred.No. 6.5e-11;
Matches 257; Conservative 0; Mismatches 224; Indels 7; Gaps 3;

QY 7 CAATTTTAGAATGGATCTTAAATTTTATGTTATAGTTCAAGTTAGTAAGTAATCCACATCCGGTACATCAAGA 66
Db |||||
QY 189 CATTTCTCTGGAAGGATCTAAGGAATTAAGTAAGTAATCCACATCCGGTACATCAAGA 248
Db |||||
QY 67 CCAAGAGCTTTGAGTCCATTCATGACATCCGTGAACCG-TGTACATGTCCTCCGATGGAC 125
Db |||||
QY 249 CCAAAATCCCTGAGCTCATTTACTATCTCTATTAAACCGGTGACATGTCAGATACACTT 308
Db |||||
QY 126 TCACTTGGTTTCAATTCGGAAAAGTTTGAAGAGTGCATAGAATATTTGATTTTGGATTCT 185
Db |||||
QY 309 TCAATTTCCCTTCATAATAAATTTGCTCAAAATTTCTCTTCGCACACATATAATTTTGACACTC 368

```

```

QY 186 TTCACTCGTTGGTGCCTTCATGAGTGAACCTCAAGAGTCCTCCAAAATATCAAAAGCCGAA 245
Db |||||
QY 369 TTCAAAATGTTAGTATTTCTTCATGAATTTCCATTTAGCTTGTATCCAAATATATCAAGCGGTT 428
Db |||||
QY 246 TCACAAATGAAATGTTGATTTGATTTTCTTGTCTAATGCACAAAACAGGGCATTCATA 305
Db |||||
QY 429 GTGAGATCTTGTATCTTATTTGAATCTCATGATGTTTGAAGCATCATAGAACACATTTCAIG 488
QY 306 GCCTTTGTTGTTTAAAGCAAAACATTTCTTCCTCGATTCATCCCATTCGCTCATCGGAAGA 365
Db |||||
QY 489 GCTTGACCAATGATTAGAACATTTTTCATTATC-----ACTAGTCAATAGTTTCATC--TTGC 542
QY 366 GAAAATTTTGAATCCATTTTCGACAATAGACAAAGCTCGAAATCCATCGAAGAAATGAGG 425
Db |||||
QY 543 TTCAATACAAACAAATTCATCTCTGSCAATTTGGCAAAACCTTTCCACTCATTTGCTCTATGA 602
QY 426 AAGATCCTCATATGAGTTTTCCTCAATACATGTAATTCGACTCATTAACATAGTGGATGT 485
Db |||||
QY 603 TGCATTCGACATCTTATTTTCTTAATAGTCATATATTTGTTCTCATCAAAATGTGATGGTTT 662
QY 486 GTAATGAA 493
Db |||||
QY 663 CCAATGCA 670
Db |||||

RESULT 14
CG802722/c
LOCUS ih34f08.b1 WGS-ZmaysU (DH10b or JM107MA2) Zea mays genomic clone
DEFINITION ih34f08, genomic survey sequence.
ACCESSION CG802722
VERSION CG802722.1 GI:32626692
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 649)
AUTHORS Rabinowicz,P.D., O'Shaughnessy,A.L., Palmer,L., Ballija,V.,
Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S.,
Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.
TITLE Genomic shotgun sequences from Zea mays (unfiltered)
JOURNAL Unpublished (2003)
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mccombie@cshl.org
Plate: ih34 row: f column: 08
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 649.
Location/Qualifiers
1..649
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/lab_host="DH10b or JM107MA2"
/clone="ih34f08"
/note="Organ: Immature ears; Site 1: Xba I; Site 2: Xba I;
The vector was digested with XbaI and one nucleotide was
added by fill in in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, b/g reads in pUC19). The ligation was
transformed into DH10b or JM107MA2."

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:23:59 ; Search time 8134.53 Seconds  
(without alignments)  
11509.077 Million cell updates/sec

Title: US-09-892-635A-45  
Perfect score: 2160  
Sequence: 1 ggatcccaacttttaggaat.....agcctctcattcaccatgg 2160

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2110	97.7	4924	6	BD080859	Banana pr
2	185.2	8.6	298	8	AF399940	Musa acum
3	122.2	5.7	282699	8	AF506028	Poncirus
4	87.8	4.1	86019	2	AC133571	Medicago
5	87.8	4.1	121952	8	AC146745	Medicago
6	83.8	3.9	192284	2	AC146910	Medicago
7	83.4	3.9	132605	8	AP004957	Lotus cor
8	83.2	3.9	141590	2	AC126788	Medicago
9	83	3.8	141590	2	AC126788	Medicago
10	81.4	3.8	125883	8	AC124959	Medicago
11	81.2	3.8	100810	8	AF411804	Lycopersi
12	79.8	3.7	106632	8	AP004500	Lotus cor
13	79.4	3.7	99606	8	AC116033	Lotus cor
14	78.8	3.6	129921	8	AC146568	Medicago
15	78.4	3.6	135983	2	AC144538	Medicago
16	78.4	3.6	136470	2	AC145222	Medicago
17	76.6	3.5	97579	2	AC146722	Medicago
18	74.4	3.4	93815	2	AC137663	Cicer ari
19	73.6	3.4	133116	2	AC146631	Medicago
20	73.4	3.4	809	6	A70187	Sequence 48
21	71.8	3.3	104684	8	AP006124	Lotus cor
22	71.6	3.3	107979	8	AP006146	Lotus cor
23	70.6	3.3	127857	2	AC133780	Medicago
24	70.4	3.3	80930	8	AP004894	Lotus cor
25	70.4	3.3	127739	2	AC145156	Medicago
26	70.2	3.2	59202	8	AF220603	Lycopersi
27	69.6	3.2	167484	2	AC145225	Zea mays
28	69	3.2	101110	8	AP006141	Lotus cor
29	69	3.2	114097	2	AC122161	Medicago
30	68.8	3.2	75051	8	AP006114	Lotus cor
31	68.8	3.2	112721	8	AP006401	Lotus cor
32	68.8	3.2	127049	8	AP006428	Lotus cor
33	68.8	3.2	129532	8	AP004482	Lotus cor
34	68	3.1	72616	8	AP006091	Lotus cor
35	68	3.1	104321	8	AP006142	Lotus cor
36	67.4	3.1	105246	8	AC139747	Medicago
37	67.4	3.1	106730	8	AC144724	Medicago
38	67.4	3.1	115995	8	AC146566	Medicago
39	67.2	3.1	79882	8	AP006118	Lotus cor
40	67.2	3.1	80018	8	AP006355	Lotus cor
41	67.2	3.1	86350	8	AP006139	Lotus cor
42	67.2	3.1	98492	8	AP006085	Lotus cor
43	67.2	3.1	100864	8	AP006391	Lotus cor
44	67.2	3.1	106033	8	AP006128	Lotus cor
45	67.2	3.1	108987	8	AP004511	Lotus cor

# ALIGNMENTS

RESULT 1  
BD080859  
LOCUS  
DEFINITION Banana proteins, DNA, and DNA regulatory elements associated with  
fruit development.  
ACCESSION BD080859  
VERSION BD080859.1 GI:22626462  
KEYWORDS JP 2001517446-A/9.  
SOURCE Musa acuminata  
ORGANISM Musa acuminata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Zingiberales; Musaceae;  
Musa.  
REFERENCE 1 (bases 1 to 4924)  
AUTHORS May, G. and Clendennen, S.

Pred. No. is the number of results predicted by chance to have a

TITLE Banana proteins, DNA, and DNA regulatory elements associated with fruit development  
JOURNAL Patent: JP 2001517446-A 9 OCT-2001;  
BOYCE THOMPSON INSTITUTE FOR PLANT RESEARCH INC  
COMMENT OS Musa acuminata (banana)  
PN JP 2001517446-A/9  
PD 09-OCT-2001  
PF 23-SEP-1998 JP 2000512960  
PR 23-SEP-1997 US 60/060062  
PI GREGORY MAY,STEPHANIE CLENDENEN  
PC C12N15/09,A01H5/00,C07K14/415,C07K14/42,C07K14/825,C12N5/10,  
PC C12N9/08,  
PC C12N9/10,C12N9/24,C12N15/00,C12N5/00  
CC Banana proteins, DNA, and DNA regulatory elements associated with fruit development  
CC development  
CC Key  
FH Key  
FT source  
FT Location/Qualifiers  
FEATURES source  
1..4924  
Location/Qualifiers  
/organism='Musa acuminata (banana)'.  
/mol\_type='genomic DNA'  
/db\_xref='taxon:4641'  
ORIGIN  
Query Match 97.7%; Score 2110; DB 6; Length 4924;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 2151; Conservative 0; Mismatches 1; Indels 9; Gaps 3;  
QY 1 GGATCCCAACTTTTAGGAATGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAA 60  
DB 1 GGATCCCAACTTTTAGGAATGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAA 60  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTGATGACATCCGTTGAAACCGTGTACATGTCCTCGA 120  
DB 61 TCTTTACCAAGAGCTTTGAGTCCATTGATGACATCCGTTGAAACCGTGTACATGTCCTCGA 120  
QY 121 TGGACTCACTGGTTTCAATTCGGAAGTTTCGAAAGTGCATAGAAATATGATTTGG 180  
DB 121 TGGACTCACTGGTTTCAATTCGGAAGTTTCGAAAGTGCATAGAAATATGATTTGG 180  
QY 181 ATTCTTTCACCTCGGTGGTCCCTCATGAGTGACCTCAAGAGTCCCTCCAAATATCAAAAG 240  
DB 181 ATTCTTTCACCTCGGTGGTCCCTCATGAGTGACCTCAAGAGTCCCTCCAAATATCAAAAG 240  
QY 241 CCGAATCAAAATGAAATGTGAATTCATTTTGTCTTAATGCAACAAACAGGGCAT 300  
DB 241 CCGAATCAAAATGAAATGTGAATTCATTTTGTCTTAATGCAACAAACAGGGCAT 300  
QY 301 TCATAGCCCTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCATCG 360  
DB 301 TCATAGCCCTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCATCG 360  
QY 361 GAAGAGAAATTTTGAATTCATTTTCGCAATAGACCAAGCTCGAAATCCATGTCATG 420  
DB 361 GAAGAGAAATTTTGAATTCATTTTCGCAATAGACCAAGCTCGAAATCCATGTCATG 416  
QY 421 GAAATGAGAGAGATCCTCATATGATGTTTCCAAATACATGTAATTCGATCAATTAACATA 480  
DB 417 GAAATGAGAGAGATCCTCATATGATGTTTCCAAATACATGTAATTCGATCAATTAACATA 476  
QY 481 GGTGATGTGAATGAATGACCTCATGCTATCTCTCTGGGTATTAAACCAATAT 540  
DB 477 GGTGATGTGAATGAATGACCTCATGCTATCTCTCTGGGTATTAAACCAATAT 536  
QY 541 GAGAGTGAGCCCTTGTCTGATACCAATTTGTAGGATCAGAGTGGCACTAAGAGAGGGGG 600  
DB 537 GAGAGTGAGCCCTTGTCTGATACCAATTTGTAGGATCAGAGTGGCACTAAGAGAGGGGG 596  
QY 601 GAGTGAATAGTGAGTGAATTAACATATTAAGTTTAAATTAATGTAATTCGTAATACGAG 660  
DB 597 GAGTGAATAGTGAGTGAATTAACATATTAAGTTTAAATTAATGTAATTCGTAATACGAG 656

QY 661 AAGATTTGTTTAAATAGTAACCTTGAGTAGATGAAACCCAAAGTTAACAGTAGTGAAA 720  
DB 657 AAGATTTGTTTAAATAGTAACCTTGAGTAGATGAAACCCAAAGTTAACAGTAGTGAAA 716  
QY 721 TAACAATTTTCGGGAAAGTAAGAACTCACACATTCAGAGAAATACCAATTTTAAAGTGGT 780  
DB 717 TAACAATTTTCGGGAAAGTAAGAACTCACACATTCAGAGAAATACCAATTTTAAAGTGGT 776  
QY 781 CGGTCAAAATGACCTACATTCACCTTGTAAGCCCTTTTCGAGAGGCTCCCAATTCAC 840  
DB 777 CGGTCAAAATGACCTACATTCACCTTGTAAGCCCTTTTCGAGAGGCTCCCAATTCAC 836  
QY 841 TAGCAATACCTTTGAAGGGGAAGGACAAATACCTCTCTTACNACCTTTTACATGGTTC 900  
DB 837 TAGCAATACCTTTGAAGGGGAAGGACAAATACCTCTCTTACNACCTTTTACATGGTTC 896  
QY 901 ATACTCTTACAAATTTTCAACGAGAAAGAGAGGTGAACATGCAAGCAATTTGAAAA 960  
DB 897 ATACTCTTACAAATTTTCAACGAGAAAGAGAGGTGAACATGCAAGCAATTTGAAAA 956  
QY 961 GACTTGCTAAAGACTTTGCTTAAGGCTTTTTCATCTATTCATCTGCTCTCAAAAGTTGTA 1020  
DB 957 GACTTGCTAAAGACTTTGCTTAAGGCTTTTTCATCTATTCATCTGCTCTCAAAAGTTGTA 1016  
QY 1021 TTCTCTGCTGAGAAATTCAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCCA 1080  
DB 1017 TTCTCTGCTGAGAAATTCAGGGGTATTTATAGACCCCAAGAGATTTAAATTTGGGCTCCA 1076  
QY 1081 AATTTGCAATCTCTCTGCGGTTCCCGAGGTTCCCGGTGCCACCGCTGTGAGTTGACA 1140  
DB 1077 AATTTGCAATCTCTCTGCGGTTCCCGAGGTTCCCGGTGCCACCGCTGTGAGTTGACA 1136  
QY 1141 CTGACAGTGTACTAGAGGTGCCACCGCGGAGACCTCTCGGTTGTTGGCGGTGCCACCGC 1200  
DB 1137 CTGACAGTGTACTAGAGGTGCCACCGCGGAGACCTCTCGGTTGTTGGCGGTGCCACCGC 1196  
QY 1201 CTAGACTTTTTCAGCTCACTGTTGGATTCCAAATCTTGACCCCAACAGTCCGAAGTCCG 1260  
DB 1197 CTAGACTTTTTCAGCTCACTGTTGGATTCCAAATCTTGACCCCAACAGTCCGAAGTCCG 1256  
QY 1261 GTCCAAATGACCCGTAACCGGATTATAGGATTAACCCCTTAATTCCTAACCCCTTAATATG 1320  
DB 1257 GTCCAAATGACCCGTAACCGGATTATAGGATTAACCCCTTAATTCCTAACCCCTTAATATG 1316  
QY 1321 CAATCTAGCAACTGAAATATAGTCTTAAGCAAGTTTAAACCGGCAAACTCGAGTCT 1380  
DB 1317 CAATCTAGCAACTGAAATATAGTCTTAAGCAAGTTTAAACCGGCAAACTCGAGTCT 1376  
QY 1381 TCTTCCGCGCATCTTTGCGCAGACTCTGATATACCTTTGATTTCTTCTAGCGGACTCC 1440  
DB 1377 TCTTCCGCGCATCTTTGCGCAGACTCTGATATACCTTTGATTTCTTCTAGCGGACTCC 1436  
QY 1441 TAGTAGGTCCTCGATCTTTGTCGCGAGTTTAGCGAGTAGCGAGAACCTTCTCGGTGATCTCC 1500  
DB 1437 TAGTAGGTCCTCGATCTTTGTCGCGAGTTTAGCGAGTAGCGGAACTTCTCGGTGATCTCC 1496  
QY 1501 GCAACCGCGCATGATCTTTGCGGAGACTTTTCGAAACTTCGCAAGTCCCGATTTCT 1560  
DB 1497 GCAACCGCGCATGATCTTTGCGGAGACTTTTCGAAACTTCGCAAGTCCCGATTTCT 1556  
QY 1561 TCTCGTTTGGTTCCGACAGCATCTTAACGAAACTTCGAGCTCTTGAATTCATCGAATCGAA 1620  
DB 1557 TCTCGTTTGGTTCCGACAGCATCTTAACGAAACTTCGAGCTCTTGAATTCATCGAATCGAA 1616  
QY 1621 CTGACTCCGCTAGGCTTGTATATTTTCAAGGTATCATAGTTAATTCCTACATCTTA 1680  
DB 1617 CTGACTCCGCTAGGCTTGTATATTTTCAAGGTATCATAGTTAATTCCTACATCTTA 1676  
QY 1681 ACTCAATATAGGATTAATTAACCATCAATGATTTTTCATCATCAAAATTCGAC 1740  
DB 1677 ACTCAATATAGGATTAATTAACCATCAATGATTTTTCATCATCAAAATTCGAC 1736





AGGFFGNPNLNGRGTIALLRDRAFFLNDKLLVGMRLCNVPSRKNLTLYSOP  
EALPPYDCELNNGRSGOAGSNGSRSSDGTETPEKGGNETIQSGLDHLHTRV  
TPRLFFTLHVSGRHLFVLTLSWFSDSGLVSKLDDPLVKVEIALNAPNCODLL  
SPSLVESRLVDVYAGMDNKIPISMSRRAQVSGESSIAPFSQKSNTPGSKTSAPAL  
LNLVNGKASGVRLTHGNNCITE"

Query Match 5.7%; Score 122.2; DB 8; Length 282699;  
Best Local Similarity 59.8%; Pred. No. 1.1e-19;  
Matches 222; Conservative 0; Mismatches 148; Indels 1; Gaps 1;  
QY 6 CCACTTTTAGGAATGATCTTAAATTTTAAAGTTTAAAGTTTAAAGTTT 65  
Db 72738 CCACTCTCTAGTAGAGATCTAATGATTTTCTGACITTTTTCGCTATTGAAAGTCTT 72679  
QY 66 ACCAAGAGCTTGTAGTCCATGATGACATCCGTGAAC-GGTGTACATGCTCCGATGGA 124  
Db 72678 TCCTAGGGCTCCTAGAGTGTTCACTATGTCGCTGAATCTAGTATACATAGAGTACACATT 72619  
QY 125 CTCACITTTGTTTCAATTCGGAAGAGTTCGAAAGAGTGCATAGAATATTGTTTGGATTC 184  
Db 72618 TTCAITTTTGTTCATTTGMAACATTCGTATTGTCGAGTGTATCTACTATTATTAGATTC 72559  
QY 185 TTTCACCTCGTGTGGTCCCTCATGAGTACCTCAAGAGTCTCCTCAATATCAAAAGCCGA 244  
Db 72558 TTTTCACCTGGTTCGTACCTTCATAGACAACTCAAGTTTGTGCCAAATCTCATTAGCACT 72499  
QY 245 ATCACAATTTGAATGTGATTGAATTCATTTTGTCTAATGTCACAAACAGGCAATTCAT 304  
Db 72498 TTGCAACTAGACACTCTATGAACCTTTACTTATCTAGTGCACAGAACAGGCAATTCAT 72439  
QY 305 AGCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCATCGGAAG 364  
Db 72438 GGCCTTGAATTTAGAGAAGCTTTTCTCTATCGAATTCATTCAAATCCCGTGAGGTTT 72379  
QY 365 AGAAATTTT 375  
Db 72378 TGAATATCTT 72368

RESULT 4  
AC133571/c  
LOCUS  
DEFINITION  
Medicago truncatula clone mth2-10h12, WORKING DRAFT SEQUENCE, 2 ordered pieces.  
AC133571  
VERSION  
AC133571.28 GI:38175857  
KEYWORDS  
HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE  
Medicago truncatula (barrel medic)  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
REFERENCE  
1 (bases 1 to 86019)  
Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE  
Medicago truncatula BAC Clone mth2-10h12  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 86019)  
Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE  
Submitted (14-SEP-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
JOURNAL  
Unpublished  
REFERENCE  
3 (bases 1 to 86019)  
Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (05-NOV-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
COMMENT  
On Nov 5, 2003 this sequence version replaced gi:38044179.

----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 23532: contig of 23532 bp in length  
\* 23533 23632: gap of unknown length  
\* 23633 86019: contig of 62387 bp in length.

## FEATURES

Location/Qualifiers  
source  
1..86019  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mth2-10h12"  
/clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Query Match 4.1%; Score 87.8; DB 2; Length 86019;  
Best Local Similarity 55.0%; Pred. No. 5.1e-11;  
Matches 193; Conservative 0; Mismatches 157; Indels 1; Gaps 1;  
QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTTAAAGTTCAAGTTAGAAAAA 60  
Db 56956 GGTCTCCATTGAGTTGGAAGACATCTTAGTATTCTTTCTCTGCGGTGAGAATAA 56897  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTCATGACATCCGTGAAACG-GGTACATCTCTCG 119  
Db 56896 GTTTTCCAGTGATCTTAGTTCAATTACTATAGATGTAATCTAGATACATTCAATCA 56837  
QY 120 ATGGAATCACTTGGTTTCATTCGGAAGAGTTTCAAGAGTGCATAAGAAATTTGATTG 179  
Db 56836 ATGTTTTCATCTTCAATTCATTTCAAGATTTTCAAACTTTCTCACCAATGTCAATCTT 56777  
QY 180 GATTCCTTCATCTCGTTGGTGGCTTCATGAGTACCTCAAGATCTCTCAAAATCAAAA 239  
Db 56776 GTTCTTTAACATGGCTTGTCTCTCATGATGAGACTTTGAGAGTATCCCAACCTTTT 56717  
QY 240 GCGAATCACAAAATTCGAAATGTGATTGAAATTTTGTCTAATGCACAAAAACAGGCA 299  
Db 56716 GCAGTATCACATTCATCACTCTTTTCACTTTCTTCCATGCTTAGTGACATGATAGAA 56657  
QY 300 TTCAATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCAT 350  
Db 56656 AGTCGAGCTTTAGAAATTTAAAGATTTTCTTTTCTTCTTTTCTTTTCTCCAT 56606

## RESULT 5

AC146745

LOCUS

DEFINITION

AC146745

ACCESSION

VERSION

AC146745.17 GI:39573788

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula (barrel medic)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

1 (bases 1 to 121952)

Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.

TITLE

Medicago truncatula BAC Clone mth2-53p19

JOURNAL

Unpublished

```

REFERENCE
AUTHORS      2 (bases 1 to 121952)
              Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
              and Roe,B.A.
TITLE
JOURNAL      Submitted (24-SEP-2003) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      3 (bases 1 to 121952)
              Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
              and Roe,B.A.
TITLE
JOURNAL      Direct Submission
              Submitted (06-DEC-2003) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE
AUTHORS      4 (bases 1 to 121952)
              Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
              and Roe,B.A.
TITLE
JOURNAL      Direct Submission
              Submitted (09-DEC-2003) Department Of Chemistry And Biochemistry,
              The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT
-----
On Dec 9, 2003 this sequence version replaced gi:39229182.
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
source
Location/Qualifiers
1..121952
   /organism="Medicago truncatula"
   /mol_type="genomic DNA"
   /db_xref="taxon:3880"
   /clone="mth2-5919"
   /clone_lib="Medicago truncatula BAC library H2"
ORIGIN
Query Match      4.1%; Score 87.8; DB 8; Length 121952;
Best Local Similarity 55.0%; Pred. No. 5.2e-11;
Matches 193; Conservative 0; Mismatches 157; Indels 1; Gaps 1;
QY 1 GATCCCCAACTTTTAGGAATGATCTTAAATTTTAGTTATAGTTCAAAGTTAGAAAA 60
Db 94945 GGTCTCCATGCTGTTGGAAGACATCTTAAATTTTCTTATTCGATCATGATGGTATAA 95004
QY 61 TCTTTACCAAGAGCTTTGAGTCATTGATGATCCGCGAAC-GGTGATCATGCTCGG 119
Db 95005 GTTTTCCAGTGATCTTAATTCATTCACATAGATGTAATCTGGAGTACATTTTCATCA 95064
QY 120 ATGGACTCACTGGTTTCATTCCGAAAAGTTCCGAAAGAGTGCATTAAGATATGATTTTG 179
Db 95065 ATGTTCTCATCTTCGTTTCATTTCGAAGATTTCAAATTTTCTTACGCCATATCTATTCTT 95124
QY 180 GATTCCTTTCACTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCTCCCAATATCAAAA 239
Db 95125 GTTTCCTTTACGTGCTAGTTCCTTCATGATGAATTTTAGAGTGTCGCAAACTTCTTTT 95184
QY 240 GCCGAATCACAATTTGAATGATGATTCATTTTGTCTAAATGCACAAACAGGGCA 299
Db 95185 GCAGTAACACATTCATCACTCTTTCACCTTTCCATGCTTAGTGGCGATGATAGAAG 95244
QY 300 TTCATAGCCTTTGTGTTTAAAGCAAAAACATTTCTTCGATTCATCCCAT 350
Db 95245 AGTCGAGCTTUGAGTTAGAGTACTTTTCTTCTTATCTTCTTTTGACCAT 95295
RESULT 6
AC146910/c
LOCUS      109284 bp DNA linear HTG 06-DEC-2003
DEFINITION Medicago truncatula clone mth2-59121, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION  AC146910
VERSION    AC146910.5 GI:39228867
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM    Medicago truncatula (barrel medic)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 109284)
  Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
  and Roe,B.A.
Medicago truncatula BAC Clone mth2-59121
Unpublished
REFERENCE
2 (bases 1 to 109284)
  Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
  and Roe,B.A.
Direct Submission
Submitted (21-OCT-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
3 (bases 1 to 109284)
  Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
  and Roe,B.A.
Direct Submission
Submitted (06-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
-----
On Dec 6, 2003 this sequence version replaced gi:38371848.
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 30237: contig of 30237 bp in length
* 30238 30337: gap of unknown length
* 30338 109284: contig of 78947 bp in length.
Location/Qualifiers
1..109284
   /organism="Medicago truncatula"
   /mol_type="genomic DNA"
   /db_xref="taxon:3880"
   /clone="mth2-59121"
   /clone_lib="Medicago truncatula BAC library H2"
Query Match      3.9%; Score 83.8; DB 2; Length 109284;
Best Local Similarity 54.5%; Pred. No. 5.3e-10;
Matches 189; Conservative 0; Mismatches 157; Indels 1; Gaps 1;
QY 7 CAACCTTTAGGAATGGAATCTTAAATTTTAGTTATAGTTCAAAGTTAGAAAAATCTTTA 66
Db 1100 CAAGATCTGGAGACATCTTAAATTTTCTTACTCTTCATGATGTAATATCCCTT 1041
QY 67 CCAAGAGCTTTGAGTCCATTCATGATGACATCCGCTGAAACGGTG-TACATGCTCCGATGGAC 125
Db 1040 TCAAGAGAGCGTAACCTCATTCATATGATAGTAAACCTTCCATACATTCATATGTT 981
QY 126 TCACCTTGTTTCATTCGGAAGATTTCGAAAGAGTGCATAAGAATATTCATTTTGGATTCCT 185
Db 980 TCGGTTCTCTTCATTCGANTAGTTCAAACTTTCGACACCAATGTCATCTTCTTCT 921
QY 186 TTCACCTCGTGGTGGTCCCTTCATGATGACCTCAAGAGTCTCCCAATATCAAAAGCCGAA 245
Db 920 TTACGTCGATGGTGGTCCCTTCATGATGATGACCTTCAATGTGTCCCAATTTCTTTGGCATTT 861
QY 246 TCACAAATTGAAATGTGATTCATTTTGTCTAATGCAACAAACAGGCGATTCATA 305

```



```

Db      860  TTGCATTGTGTACTTATCATATATCTTCCTCGTCAATGCCAGTGTTAGAAATAGTTTA 801
QY      306  GCCTTTGTGTTAAAGCAAAACATCTTCTCCGATTCATCCCATTC 352
Db      800  GCCTTAAATTAATAACATGCGCTTACCTTCTTCCTGACCAGTC 754

RESULT 7
LOCUS   AP004957               132605 bp    DNA        linear    PLN 22-JUL-2003
DEFINITION
Lotus corniculatus var. japonicus genomic DNA, chromosome 2,
AP004957
ACCESSION
AP004957.1  GI:21907975
VERSION
HTG.
KEYWORDS
SOURCE   Lotus corniculatus var. japonicus (Lotus japonicus)
ORGANISM
Lotus corniculatus var. japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;
Lotus.
REFERENCE
1  Kaneo, T., Nakamura, Y., Asamizu, E., Kato, T., Sato, S. and Tabata, S.
   Structural Analysis of a Lotus japonicus Genome. I. Sequence
   Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb
   Regions of the Genome
   Unpublished
JOURNAL
REFERENCE
2  (bases 1 to 132605)
AUTHORS
Nakamura, Y.
TITLE
Direct Submission
JOURNAL
Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: yna@kazusa.or.jp,
URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
FEATURES
     source
     1. .132605
        /organism="Lotus corniculatus var. japonicus"
        /mol_type="genomic DNA"
        /variety="japonicus"
        /db_xref="taxon:34305"
        /chromosome="2"
        /clone="LjT34H20"
        /clone_lib="LjT library"
        /note="TAC clone: TW0134-synonym: Lotus japonicus"

ORIGIN
Query Match      3.9%; Score 83.4; DB 8; Length 132605;
Best Local Similarity 53.7%; Pred. No. 6.7e-10;
Matches 194; Conservative 0; Mismatches 166; Indels 1; Gaps 1;

QY      1  GGATCCCACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAGTTAGAAAAA 60
Db      112738  GGATCCAGCTTCAGGAAGACCTCTAAGAATCCTCATAACATGATCACCAGTTGTGTAA 112797

QY      61  TCTTTACCAAGAGCTTTGAGTCCCATGATGA-CATCGTGAAACGGTGATCATGTCTCG 119
Db      112798  CTTTATTAGGGGTCTAATTCAGCAATGAGTAAGTAATCTGAGAACATGCTCTCA 112857

QY      120  ATGGAATCACTGGTTTCATTCGGAAAAGTTGGAAGAGTGCATAGAATAATGATTG 179
Db      112858  ATGGAATCACTGGATTCATTCGGAAGGATTCTACTTCTGATCAAGGACACGCTTT 112917

QY      180  GATTCCTTCACTGGTTGGTCCCTTATGATGACCTCAAGATCCTCCAAATATCAAA 239
Db      112918  GACTCTTTCACCTCTGTGTTTCATCATGGGTCTCTTCAAGAAATCAAGATACCCCTTA 112977

QY      240  GCCGAATCAAAATGAAATGTGATTGAATTCATTTTTTGTCTAAATGCACAAACAGGCA 299
Db      112978  GCATATCTCAGATCTGTAATCTTCTGTTATCTTCAATTAAGAAATAGACTGGAAGAATA 113037

QY      300  TTCATAGCCTTTGTGTTTAAAGCAAAACATCTTCTCCGATTCATCCCATTCGCTCATC 359

```

```

Db      113038  GTTCTAGCTTTATGATGCAGTCAGTAAGCTTCTTCTGCTTCGATCACTCTGTTTG 113097
QY      360  G 360
Db      113098  G 113098

RESULT 8
LOCUS   AC126788/c             141590 bp    DNA        linear    HTG 11-NOV-2003
DEFINITION
Medicago truncatula clone mth2-33121, WORKING DRAFT SEQUENCE, 8
unordered pieces.
ACCESSION
AC126788
VERSION
GI:38257184
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Medicago truncatula (barrel medic)
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1  (bases 1 to 141590)
AUTHORS
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE
Medicago truncatula BAC Clone mth2-33121
JOURNAL
Unpublished
REFERENCE
2  (bases 1 to 141590)
AUTHORS
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE
Direct Submission
JOURNAL
Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
3  (bases 1 to 141590)
AUTHORS
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
TITLE
Direct Submission
JOURNAL
Submitted (11-NOV-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Nov 11, 2003 this sequence version replaced gi:37059850.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4380: contig of 4380 bp in length
* 4381 4480: gap of unknown length
* 4481 14258: contig of 9778 bp in length
* 14259 14358: gap of unknown length
* 14359 20369: contig of 6311 bp in length
* 20370 20769: gap of unknown length
* 20770 25083: contig of 4314 bp in length
* 25084 25183: gap of unknown length
* 25184 41406: contig of 16223 bp in length
* 41407 41506: gap of unknown length
* 41507 60351: contig of 18845 bp in length
* 60352 60451: gap of unknown length
* 60452 87167: contig of 26716 bp in length
* 87168 87267: gap of unknown length
* 87268 141590: contig of 54323 bp in length.
Location/Qualifiers
1. .141590
/organism="Medicago truncatula"
/mol_type="genomic DNA"

```

```
/db_xref="taxon:3880"
/clone="mth2-33i21"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match      3.9%; Score 83.2; DB 2; Length 141590;
Best Local Similarity 54.3%; Pred. No. 7.5e-10;
Matches 189; Conservative 0; Mismatches 158; Indels 1; Gaps 1;

Qy 6 CCAACTTTTGAAGATGATCTTAAATTTTATAGTTTCAAGTTCAAGTTAGAAAAATCTTT 65
Db 83904 CCAAGATCTTGGGAGACATCTTAAATTTTCTTACTCTTTCATGGATGTGAAAATCCTT 83845

Qy 66 ACBAAGAGCTTTGAGTCATTCATGATGACATCCGCGAAGGGTG-TACATGCTCCGATGGA 124
Db 83844 TTCAAGAGAGCGTAACCTCATCATATGATAGTAACCTTCCATACATTCATCATGGT 83785

Qy 125 CTCACCTTGGTTTCATTCGAAAAAGTTCGAAAGAGTGCATAGAATATGATTTGGATTC 184
Db 83784 TTCCGTTCTTTCATTCGATAGTTTCAAACTTTCGACACCAATGTCATTTCTGTTTC 83725

Qy 185 TTTCACCTGGTGTGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCGGA 244
Db 83724 TTTTACGAGTATGTTGGCTTCATGATGACGTTTAAATGATATCCCAAAATTTCTTGGCAT 83665

Qy 245 ATCAAAATGAAATGATGAATTCATTTTGTCTAATGCAACAAACAGGCAATTCAT 304
Db 83664 TTTGCATCTTGTACTCTATCATATTTCTTCCCTGCTCAATGCCATGTAGAAATAGTTT 83605

Qy 305 AGCCTTTGTGTTAAAGCAAAACATTTCTTCCGATTCATCCCATTC 352
Db 83604 AGCTTTAAATTTAATGATGCGCTTACCTTCTTCTGACCAGTC 83557

RESULT 9
AC126788 LOCUS
DEFINITION Medicago truncatula clone mth2-33i21, WORKING DRAFT SEQUENCE, 8
unordered pieces.
ACCESSION AC126788
VERSION AC126788.9 GI:38257184
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 141590)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,B., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-33i21
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 141590)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 141590)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Nov 11, 2003 this sequence version replaced gi:37059850.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4380: contig of 4380 bp in length
* 4381 4480: gap of unknown length
* 14258: contig of 9778 bp in length
* 14358: gap of unknown length
* 14359 20669: contig of 6311 bp in length
* 20670 20769: gap of unknown length
* 20770 25083: contig of 4314 bp in length
* 25084 25183: gap of unknown length
* 25184 41408: contig of 16223 bp in length
* 41409 41507: gap of unknown length
* 41508 60351: contig of 18845 bp in length
* 60352 87167: contig of 26716 bp in length
* 87168 87267: gap of unknown length
* 87268 141590: contig of 54323 bp in length.
*
* Location/Qualifiers
* 1..141590
* /organism="Medicago truncatula"
* /mol_type="genomic DNA"
* /db_xref="taxon:3880"
* /clone="mth2-33i21"
* /clone_lib="Medicago truncatula BAC library H2"

ORIGIN
Query Match      3.8%; Score 83; DB 2; Length 141590;
Best Local Similarity 54.1%; Pred. No. 8.4e-10;
Matches 190; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

Qy 1 GGATCCCAACCTTTTAGGAATGGATCTTAAATTTTATAGTTTAAAGTTCAAGTTAGAAAA 60
Db 40054 GGTCTCCATGTGCTTGAAGACATCTTAAGTTTTTCTTATCCCATCATGATGATGATGTA 40113

Qy 61 TCCTTACCAAGAGCTTTTGAAGTCCATTCATGATGATGATGATGATGATGATGATGATG 119
Db 40114 GTTTTCCAGTGTATCTTAATTCATTCACATAGATGATGATGATGATGATGATGATG 40173

Qy 120 ATGACACTCAGTCTGTTTCAATTCGAAAGTTTGAAGAGTGCATAGAATATGATTTTG 179
Db 40174 ATGTTCTCATCTTCGTTTCATTTCAAGATTTTCAAAATTTTCTTACGCCATGCTATT 40233

Qy 180 GATTTCTTCACTCGGTTGGTGGCTTTCATGATGATGATGATGATGATGATGATGATGAT 239
Db 40234 GTTTCTTTTACATGGCTAGTTCTTTCATGATGATGATGATGATGATGATGATGATGAT 40293

Qy 240 GCCGAATCACAAATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299
Db 40294 GCAGTCACATTCATCAACTCTTTTCACTTTCTTCCATGCTTAGTGGCATGATAGAA 40353

Qy 300 TTCATAGCTTTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCAT 350
Db 40354 AGTCGAGCTTTAGAGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 40404

RESULT 10
AC124959/c LOCUS
DEFINITION Medicago truncatula clone mth2-23111, complete sequence.
ACCESSION AC124959
VERSION AC124959.11 GI:28850030
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
```

Medicago.  
 1 (bases 1 to 125883)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Medicago truncatula BAC Clone mth2-23111  
 Unpublished  
 2 (bases 1 to 125883)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission  
 Submitted (06-JUN-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 3 (bases 1 to 125883)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission  
 Submitted (05-NOV-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 4 (bases 1 to 125883)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission  
 Submitted (05-MAR-2003) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 On Mar 5, 2003 this sequence version replaced gi:24635962.  
 -----  
 Center: Genome Center  
 Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR

# FEATURES

source  
 1..125883  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-23111"  
 /clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Query Match 3.8%; Score 81.4; DB 8; Length 125883;  
 Best Local Similarity 53.5%; Pred. No. 2.1e-09;  
 Matches 192; Conservative 0; Mismatches 166; Indels 1; Gaps 1;  
 1 GGATCCCACTTTAGGAATGGATCTTAAATTTTGTATTAAGTTCAGAGTTAGAAAAA 60  
 5238 GGTCTCCATACACTTGGGAAGACATCTTAATATTTTCTGATTCATCATAGTAGAATA 5179  
 61 TCTTTACCAAGAGCTTTGAGTCCATTGATGACATCGGTGAACG-GTGTACATGCTCCG 119  
 5178 GCCTTTCCAGAGATCTCATTTCTATTCTATTGTGTAGTAATCTAGCATACATCTCATCA 5119  
 120 ATGGACTCATCTGGTTTCATTCGGAAGAAAGTTCGAAAGAGTGCATAAGAAATATTGTTTG 179  
 5118 ATGGTTTCATTTTCACTCATTTTCAAAAACCTTCAAAATTTTCTGACTCCAAATATCTTA 5059  
 180 GATCTTTTCACTGGTTGGTCCCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAA 239  
 5058 GTTCTTTTGACATGAGATGTACCTCATGATATATTTTTCAGAGTGCCCATACCTCTTG 4999  
 240 GCCGAATCAAAATGGAATGTGATTGATTTATTTTGTCTAATTCACAAACAGGGCA 299  
 4998 GCATTTTACACTCACTCACTCTTTCACCTTCTCCATGGTTAGTGACATGATAGAAA 4939  
 300 TTCATAGCCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCAT 358  
 4938 AGCTAGCTTTAGAGTTTAGTATAATACATGAGCCTTTTCACTGTTGACCATGCTTCTT 4880

## RESULT 11

AF411804

LOCUS  
 DEFINITION  
 AF411804 Lycopersicon esculentum BAC clone Clemson\_ID 47113, complete sequence.  
 ACCESSION  
 VERSION  
 AF411804.1 GI:15987769  
 KEYWORDS  
 SOURCE  
 Lycopersicon esculentum (tomato)  
 ORGANISM  
 Lycopersicon esculentum  
 Sukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
 REFERENCE  
 1 (bases 1 to 100810)  
 Van Der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and  
 Tanksley, S.  
 Deductions about the Number, Organization, and Evolution of Genes  
 in the Tomato Genome Based on Analysis of a Large Expressed  
 Sequence Tag Collection and Selective Genomic Sequencing  
 Plant Cell 14 (7), 1441-1456 (2002)  
 12119366  
 REFERENCE  
 2 (bases 1 to 100810)  
 van der Hoeven, R.S. and Tanksley, S.D.  
 Direct Submission  
 Submitted (20-AUG-2001) Plant Breeding, Cornell University, 252  
 Emerson Hall, Ithaca, NY 14850, USA  
 FEATURES  
 source  
 1..100810  
 /organism="Lycopersicon esculentum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:4081"  
 /clone="Clemson\_ID 47113 BAC"  
 ORIGIN  
 Query Match 3.8%; Score 81.2; DB 8; Length 100810;  
 Best Local Similarity 52.0%; Pred. No. 2.4e-09;  
 Matches 205; Conservative 0; Mismatches 188; Indels 1; Gaps 1;  
 4 TCCCAACTTTTAGGAATGGATCTTAAATTTTGTATTAAGTTCAGAGTTAGAAAAATCT 63  
 89844 TCCCAAGACCTTAGAAAAATCTTAAGTACTTTGTCAACCTGTTCTCAGTGGTATATACT 89903  
 64 TTACCAAGAGCTTTGAGTCCATTGATGACATCGTGAAAC-GGTGTACATGCTCCGATG 122  
 89904 TTACCAATGAGATTAGTTCGTTTATCAGAGTGTAATCTAGTATCATATCTTAGA 89963  
 123 GACTCATCTGGTTTCATTCGGAAGAAAGTTCGAAAGAGTGCATAAGAAATATTGATTTGGAT 182  
 89964 GATTTTCCGGAATTCATCTTAAAGGCTTCATATTTCTGAACACAACTAGCAATCTTGAAT 90023  
 183 TCTTTCACTCGGTTGGTTCCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCC 242  
 90024 TTTCCGACTTGAGTTGTACCTTCATGAGAGTTTGTAAAGTATCCCATATTTGTTAGCA 90083  
 243 GAATCAAAATGAAATGTGATTGAATTCATTTTGTCTAATGCACAAACAGGGCATTC 302  
 90084 GAGGTAACTTGAGATACGGTTAACTCATCGGTCCTAGTCCCATACAGAAATGCAC 90143  
 303 ATAGCCTTTGTGTTTAAACAAACAAATTCCTTCGAGATTCATCCCATTCGCTCATCGGA 362  
 90144 TTAGCTTTTGCATCTCTTCCCTAGCATTTTATAATCGCTTCATTTATATGATCTGCAGGT 90203  
 363 AGAGAAAATTTTGAATCCATTTTCGACAATAG 396  
 90204 TTAGGTACTTTTATTTCCCTTCACATCATCAATAG 90237  
 RESULT 12  
 AP004500/c  
 LOCUS  
 DEFINITION  
 AP004500 Lotus corniculatus var. japonicus genomic DNA, chromosome 1,  
 clone:LjT05P21, TM0032, complete sequence.  
 ACCESSION  
 VERSION  
 AP004500.1 GI:17736867  
 KEYWORDS  
 HTG.

SOURCE Lotus corniculatus var. japonicus (Lotus japonicus)  
ORGANISM Lotus corniculatus var. japonicus  
TITLE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteae;  
Lotus.  
REFERENCE 1  
AUTHORS Sato,S., Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T. and Tabata,S.  
TITLE Structural Analysis of a Lotus japonicus Genome. I. Sequence  
Features and Mapping of Fifty-six TAC clones which cover the 5.4 Mb  
Regions of the Genome  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 106632)  
Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (13-DEC-2001) Yasukazu Nakamura, Kazusa DNA Research  
Institute, Department of Plant Gene Research; 1532-3, Yana,  
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,  
URL:http://www.kazusa.or.jp, Tel:81-438-52-3935,  
Fax:81-438-52-3934)  
FEATURES  
source  
location/Qualifiers  
1..106632  
/organism="Lotus corniculatus var. japonicus"  
/mol\_type="genomic DNA"  
/variety="japonicus"  
/db\_xref="taxon:34305"  
/chromosome="1"  
/clone="LJ705P21"  
/clone\_lib="LJ7 library"  
/note="TAC clone:TM0032-synonym: Lotus japonicus"  
ORIGIN  
Query Match 3.7%; Score 79.8; DB 8; Length 106632;  
Best Local Similarity 53.9%; Pred. No. 5.4e-09;  
Matches 185; Conservative 0; Mismatches 157; Indels 1; Gaps 1;  
QY 11 TTTTAGGAATGATCTTAAATTTTGTATTAAGTCAAGTTCAGAAATCTTTACCA 70  
Db TCTGGGAGAGATCTCAAGATCTTCTTCAAGCTTCTCATCTGACATAGTTCTCCA 24508  
QY 71 GAGCTTTGAGTCCATGATGATCATCCGTGAAACG-GTGTACATGTCTCCGATGGACTCAC 129  
Db AAGCAAAATGAAGCATTTAGACAGTACCGACGACGATGTGGAATCTGAGATAGTCTTT 24448  
QY 130 TTGGTTTCATTCGGGAAAAGTTCGAAAGATGCAATGAATATTTGATTTGGATCTTTCA 189  
Db CTTTAGTCATCATCAAAATTTTCAAAATTTGAGTGGTGAAGAACTGGAAGTCTTGACATCTCTA 24388  
QY 190 CTCGGTTGGTGCCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCCGAATCAC 249  
Db CTCGAGTAGTTCTTCATGTGAGTCTTCAGAAATTTCCAGATATCTTTGCCACAGTAC 24328  
QY 250 AAATTTGAATGTGATGAATTCATTTTGTCTTAATGCAACAAACAGGGCATTCATAGCCT 309  
Db ACGTATTGATCAGCTAAACATATTTCTTGTCACTCCATTAAGATAGATTCAGAGCT 24268  
QY 310 TTGTGTTTAAGCAAAAACATCTTCTCCGATTCATCCCATTC 352  
Db TTGAGTTTCCAAGACAGCTTCACTTCAACACTGGTCCATTC 24225  
RESULT 13  
AC116033/c 99606 bp DNA linear PLN 04-OCT-2003  
LOCUS  
DEFINITION Zea mays genomic clone ZM15C05 sequence, complete sequence.  
ACCESSION AC116033  
VERSION AC116033.3 GI:37514986  
KEYWORDS HTG  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 99606)  
Nagaki,K., Song,J., Stupar,R.M., Parokony,A.S., Yuan,Q.,  
Yuang,S., Liu,J., Hsiao,J., Jones,K.M., Dawe,K., Buell,R. and  
Jiang,J.  
TITLE Molecular and cytological analyses of large tracks of centromeric  
DNA reveal the structure and evolutionary dynamics of maize  
centromeres  
Genetics 163 (2), 759-770 (2003)  
MEDLINE 22505408  
PUBMED 12618412  
REFERENCE 2 (bases 1 to 99606)  
Jiang,J., Nagaki,K., Yuan,Q., Vanaken,S., Utterbach,T.,  
Gansberger,K. and Buell,R.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-2002) University of Wisconsin, Department of  
Horticulture, Madison, WI 53706, USA  
REFERENCE 3 (bases 1 to 99606)  
Jiang,J.  
TITLE Direct Submission  
JOURNAL Submitted (21-NOV-2002) University of Wisconsin, Department of  
Horticulture, Madison, WI 53706, USA  
REFERENCE 4 (bases 1 to 99606)  
Jiang,J.  
TITLE Direct Submission  
JOURNAL Submitted (28-NOV-2002) University of Wisconsin, Department of  
Horticulture, Madison, WI 53706, USA  
REFERENCE 5 (bases 1 to 99606)  
Jiang,J.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-2002) University of Wisconsin, Department of  
Horticulture, Madison, WI 53706, USA  
REFERENCE 6 (bases 1 to 99606)  
Jiang,J.  
TITLE Direct Submission  
JOURNAL Submitted (04-OCT-2003) University of Wisconsin, Department of  
Horticulture, Madison, WI 53706, USA  
COMMENT On Oct 4, 2003 this sequence version replaced gi:25141017.  
FEATURES  
source  
location/Qualifiers  
1..99606  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4577"  
/chromosome="unknown"  
/clone="ZM15C05"  
/note="The BAC contains DNA sequences associated with  
maize centromeres."  
ORIGIN  
Query Match 3.7%; Score 79.4; DB 8; Length 99606;  
Best Local Similarity 51.9%; Pred. No. 6.8e-09;  
Matches 256; Conservative 0; Mismatches 226; Indels 11; Gaps 3;  
QY 4 TCCCAACTTTAGGAATGGATCTTAAATTTTGTATTAAGTTCAAAGTTAGAAAATCT 63  
Db TCATATTTCTTCGGAAGTGAATCTAAGAACTGTGTGTGAATCCACATCCGGTACATTA 33064  
QY 64 TTACCAAGAGCTTTGAGTCCATGATGACATCCGTGAACGCT-GTACATGCTCTCCGATG 122  
Db AACCAAGTCTTTGAGTTCATTTATCTCTCATTCACCAAGTGAACATATCGGATACA 33004  
QY 123 GACTCAGTCTGGTTTCATTCGGAAAAGTTTCGAAAGAGTGCATAGAATAATTCATTTTGGAT 182  
Db CTCATCTCTCTTTCATATAAATTTGCTCAAACTTCCCTTTGCACACATATAGTTTGCA 32944  
QY 183 TCTTTCACTCGTTGGTGGCTTCATGAGTGCCTCAAGAGTCCCTCCCAATATCAAAAGCC 242  
Db CTCCTCAATTTGAGTTCCTTCATGAATTTCCATTAAGTATTTATCCAAATCTCATGAGCG 32884  
QY 243 GAATCACAAATTGAAATGTGATTCGAATTCATTTTGTCTAATGCACAAAACAGGGCATTC 302  
Db GTTGTGAGATTTCTTGATACGATTAACCAATTTATATCAAGGATCATATAAGACATTC 32824  
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCAATTCGCTCATCGGA 362

Db 32823 TTGGCTTGATCATTTGTGAGGATATTCATTAATACATCAAGATGGTTTCATCTCTC 32764  
 QY 363 AGAGAAATTTTGAATCCATTTTCGACATAGACAAAGCTGAAATCCATGATGGA 422  
 Db 32763 AACACAA-----CAAAATCCATCCCTGACAAATGGCCAAATTT-----TTCCACCCATTCG 32714  
 QY 423 AATGAGGAAGATCCTCATATGAGTTTTCATATACATGTAATTCGACTCATTAACATAGG 482  
 Db 32713 TCTAAGATGATTCACATCTTAATTTTCCAAATATATCATTAATTTTCCATCAAAAGTGGCG 32654  
 QY 483 TGGATGTGTAATG 495  
 Db 32653 TGGCTTTCCAATG 32641

RESULT 14  
 AC146568 129921 bp DNA linear PLN 18-DEC-2003  
 LOCUS Medicago truncatula clone mth2-101018, complete sequence.  
 DEFINITION  
 AC146568  
 VERSION  
 AC146568.11 GI:38708056  
 KEYWORDS  
 HTG;  
 SOURCE  
 Medicago truncatula (barrel medic)  
 ORGANISM  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 129921)  
 AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
 and Roe, B.A.  
 TITLE Medicago truncatula BAC Clone mth2-101018  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 129921)  
 AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
 and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-AUG-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 3 (bases 1 to 129921)  
 AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
 and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 4 (bases 1 to 129921)  
 AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
 and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 5 (bases 1 to 129921)  
 AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.  
 and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

COMMENT On Dec 5, 2003 this sequence version replaced gi:38564350.  
 ----- Genome Center  
 Center: Department of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----

FEATURES  
 source Location/Qualifiers  
 1. 129921  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-101018"  
 /clone\_lib="Medicago truncatula BAC library H2"

ORIGIN  
 Query Match 3.6%; Score 78.8; DB 8; Length 129921;  
 Best Local Similarity 53.4%; Pred. No. 9.7e-09;  
 Matches 187; Conservative 0; Mismatches 162; Indels 1; Gaps 1;

QY 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAAA 60  
 Db 50577 GGCTGCCAATCCCTTGTCAATACATCTCAAAATTTTATAAGTGAATCTTGAATG 50636  
 QY 61 TCTTTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTTGAAC-GGTGTACATGTCCTCG 119  
 Db 50637 TCATTACCTAGATTTCGAATTTTAAACAATATGGGTGAATCTCATTTGCACGAGGAG 50696  
 QY 120 ATGGAATCCTTTGTTTCATTCGGAAAAAGTTCCGAAGAGTGCATAAGAATATTGATTTG 179  
 Db 50697 ATGGATCCACTCTCCATCGGAAGAGTTTATATTGTTGAACAAGATATTGATCAATT 50756  
 QY 180 GATTCCTTCACTCGTTGGTGGCTTCATGAGTACCTCAAGAGTCCCTCAAAATATCAAAA 239  
 Db 50757 GATTTGTTGACGTTGTTCAITTCCTTCATGAAGGTTTCCAAAGGTTCCCATAGCCTTA 50816  
 QY 240 GCGGAATCACAATTTGAAATGTGATTGAATTCATTTTGTCTAATGCACAAAACAGGCA 299  
 Db 50817 GAGTTTTCGAATGAGATACGAATGTCTCATCTCCAGAGAGGAATGAGAATG 50876  
 QY 300 TTCATAGCCTTTGTGTTTAAAGCAAAAACATTTCTCTCCGATTCATCCCA 349  
 Db 50877 TTTCGCGCTTCAATCAATCAATTGAACCTTCCTTTTGTCAATCACTCCCA 50926

RESULT 15  
 AC144538 135983 bp DNA linear HTG 05-DEC-2003  
 LOCUS Medicago truncatula clone mth2-31121, WORKING DRAFT SEQUENCE, 3  
 DEFINITION  
 ordered pieces.

ACCESSION  
 AC144538  
 VERSION  
 AC144538.11 GI:38708043  
 KEYWORDS  
 HTG; HTGS PHASE2; HTGS DRAFT.  
 SOURCE  
 Medicago truncatula (barrel medic)  
 ORGANISM  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 135983)  
 AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 TITLE Medicago truncatula BAC Clone mth2-31121  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 135983)  
 AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-APR-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 3 (bases 1 to 135983)  
 AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 4 (bases 1 to 135983)  
 AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

COMMENT On Dec 5, 2003 this sequence version replaced gi:38678558.  
 ----- Genome Center  
 Center: Department of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
\* 1 104469: contig of 104469 bp in length  
\* 104470 104569: gap of unknown length  
\* 104570 113132: contig of 8563 bp in length  
\* 113133 113232: gap of unknown length  
\* 113233 135983: contig of 22751 bp in length.

FEATURES  
    Location/Qualifiers  
    1..135983  
        /organism="Medicago truncatula"  
        /mol\_type="genomic DNA"  
        /db\_xref="taxon:3880"  
        /clone="mth2-31121"  
        /clone\_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match 3.6%; Score 78.4; DB 2; Length 135983;  
Best Local Similarity 53.4%; Pred. No. 1.2e-08;  
Matches 186; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
  
Qy 6 CCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAGTTAGAAAATCTTT 65  
Db 114529 CCAAGATCTGGGAGACATCTCAAATTTTCTTACTCTTTCATGGATTGAAAATCCTT 114470  
  
Qy 66 ACCAAGAGCTTTGAGTCCATTGATGACATCGGTGAAACGGT-GTACATGTCTCCGATGGA 124  
Db 114469 TTCAGAGAGCGTAACTCATTCTATAATGATAGTAAACCTTCCATACATTTTCATCTATGGT 114410  
  
Qy 125 CTCACCTGGTTTCATTCGGAAGAAAGTTGGAAGAGTGCATAGATATGATTTGGATTC 184  
Db 114409 TTCGGTTTCTTTTCATTTCGAATAGTTCAAACCTTTCGAACACCAATGTCAATTTCTTTTC 114350  
  
Qy 185 TTTCACCTCGGTGGTGCCTTCATGAGTGCACCTCAAGCTCCCTCCAAATATCAAAGCCGA 244  
Db 114349 CTTTACGTGACGGTGGCTTCATGATGATACCTTTAATGTATCCCAAATTTCTTGGCGTT 114290  
  
Qy 245 ATCACAATTTGAAATGTGATTTGAATTTTGTCTAATGCACAAAACAGGCGATTTCAT 304  
Db 114289 TTTGCATTTCTGTACTCTATCATATTTCTCCCTGCTCAACGCCATGTTAGAAATAGTTT 114230  
  
Qy 305 AGCCTTTGTGTTAAAGCAAAACATTTCTTCCGATTCATCCCATTC 352  
Db 114229 AGCTTTAAATTTAATAACATCGGCTTACCTTTCTTCTTGACCCAGTC 114182

Search completed: September 18, 2004, 23:51:57  
Job time : 8138.53 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:18:43 ; Search time 762.706 Seconds  
(without alignments)  
12030.986 Million cell updates/sec

Title: US-09-892-635A-45

Perfect score: 2160

Sequence: 1 ggatcccaacttttaggaat.....agctctctaccatgg 2160

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_29Jan04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002s:\*
- 7: Geneseqn2003as:\*
- 8: Geneseqn2003bs:\*
- 9: Geneseqn2003cs:\*
- 10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2105.2	97.5	4924	2	AAX25614 Banana ri
2	73.4	3.4	809	2	Aav28690 Ripening
3	62.8	2.9	2417	2	Aav22222 SIRE-1 CD
4	54	2.5	110000	3	Continuation (3 of
5	53.4	2.5	110000	3	Continuation (4 of
6	52.6	2.4	515	2	Aaq81475 HMG2 prom
7	51.4	2.4	2000	7	Ada71938 Rice gene
8	46.8	2.2	16439	6	Ab132886 Human imm
9	46.2	2.1	2000	7	Ada71938 Rice gene
10	43.4	2.0	742	2	Aav28672 Ripening
11	42.4	2.0	1186	2	Aax25608 Banana ri
12	41.6	1.9	5945	6	Ab132084 Human imm
13	41.6	1.9	7001	6	Abk33919 Human DNA
14	41.6	1.9	7001	7	Ada20394 Prostate
15	41.6	1.9	7001	7	Ada84201 Human ren
16	41.2	1.9	1722	7	AcF73552 Staphyloc
17	41.2	1.9	2881	2	Aav74565 Staphyloc
18	41.2	1.9	5879	6	Ab132269 Human imm
19	41.2	1.9	8197	6	Ab134515 Human met
20	41.2	1.9	8197	6	Ab170542 Chemocall
21	41	1.9	6022	4	Aas46662 Tumour su
22	40.8	1.9	7351	6	Ab132028 Human imm
23	40.8	1.9	18967	4	Ab123560 Drosophil

24	40.6	1.9	726	6	ABQ68510	Listeria
25	40.6	1.9	1337	6	ABQ70519	Listeria
c 26	40.6	1.9	6216	6	ABK39932	Human che
c 27	40.6	1.9	6216	6	ABL70139	Chemical
c 28	40.4	1.9	6161	6	ABL32623	Human imm
c 29	40.4	1.9	6237	6	ABL32925	Human imm
c 30	40.2	1.9	15399	6	ABL33515	Human imm
c 31	40	1.9	9483	6	ABL32377	Human imm
c 32	40	1.9	9483	6	ABL70516	Chemical
c 33	40	1.9	9483	6	AAS61092	Human gen
c 34	39.6	1.8	1722	7	ACA19762	Prokaryot
c 35	39.6	1.8	83391	6	ABQ67093	Human ang
c 36	39.4	1.8	9504	6	ABK28407	Human ang
c 37	39.2	1.8	18679	6	ABN80300	DNA trans
c 38	39	1.8	5815	6	ABL33234	Human imm
c 39	39	1.8	5987	6	ABL33562	Human imm
c 40	39	1.8	5987	6	ABQ67101	Human ang
c 41	39	1.8	5987	9	ADB54307	Pretrate
c 42	39	1.8	5987	9	ADB54179	Pretrate
c 43	39	1.8	6089	4	AAS46579	Tumour su
c 44	39	1.8	6089	6	ABL33836	Human imm
c 45	39	1.8	12025	6	ABL33299	Human imm

## ALIGNMENTS

RESULT 1

AAX25614

ID AAX25614 standard; DNA; 4924 BP.

XX AC AAX25614;

XX AC AAX25614;

DT 02-AUG-1999 (first entry)

XX XX Banana ripening fruit chitinase DNA.

DE DE Banana; fruit ripening; differential expression; fruit development;

DE DE transgenic plant; chitinase; ss.

XX XX Musa acuminata.

XX XX WO9915668-A2.

XX XX 01-APR-1999.

XX XX 23-SEP-1998; 98WO-US003343.

XX XX 25-SEP-1997; 97US-0060062P.

XX XX (BOYC-) BOYCE THOMPSON INST PLANT RES.

XX XX May G, Clendennen S;

XX XX WPI; 1999-244425/20.

XX XX P-PSDB; AAY05845, AAY05846, AAY05847.

XX XX New isolated banana DNA molecules.

XX XX Claim 5; Fig 17A-G; 143pp; English.

XX XX This is the nucleotide sequence of a chitinase DNA molecule. The

invention provides isolated DNA molecules which are differentially expressed during banana fruit development, and the protein products of these genes. The DNA is selected from a group comprising starch synthase, chitinase, endochitinase, beta-1,3-glucanase, thaumatin-like protein, ascorbate peroxidase, metallothionein, lectin and senescence-related protein. The regulatory elements of the genes can be used to produce chimeric genes for transformation of plants to provide controlled expression of heterologous DNA during fruit development, or in response to exogenous developmental signals, such as ethylene signals. The heterologous protein, e.g. a therapeutic protein, can be isolated from the fruit or consumed directly in the transformed fruit. The identity of

CC the present sequence is not given. The entire chitinase DNA sequence has  
CC been decoded in all 3 reading frames to provide the amino acid sequences  
CC given in AAY05845-47

XX  
SQ Sequence 4924 BP; 1304 A; 1154 C; 1037 G; 1425 T; 0 U; 4 Other;

Query Match 97.5%; Score 2105.2; DB 2; Length 4924;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 2148; Conservative 0; Mismatches 4; Indels 9; Gaps 3;

Qy	1	GGATCCCAACCTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAAGTTAGAAATA	60
Db	1	GGATCCCAACCTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAAGTTAGAAATA	60
Qy	61	TCCTTACCAAGAGCTTTGAGTCCATGTATGATGACATCCGGTGAACCGGTGTACATGCTCCGA	120
Db	61	TCCTTACCAAGAGCTTTGAGTCCATGTATGATGACATCCGGTGAACCGGTGTACATGCTCCGA	120
Qy	121	TGGACTCACTGGTTTCATTTCGGAAAGTTTCGAAAGAGTGCATAGAAATATTGATTTGG	180
Db	121	TGGACTCACTGGTTTCATTTCGGAAAGTTTCGAAAGAGTGCATAGAAATATTGATTTGG	180
Qy	181	ATCTTTTCACTCGGTGGTGCCTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAG	240
Db	181	ATCTTTTCACTCGGTGGTGCCTTCATGAGTGACCTCAAGAGTCTCTCCAAATATCAAAAG	240
Qy	241	CCGAATCAAAATGGAATGATGAAATTCATTTTCTTAATGTCACAAAACAGGGCAT	300
Db	241	CCGAATCAAAATGGAATGATGAAATTCATTTTCTTAATGTCACAAAACAGGGCAT	300
Qy	301	TCATAGACCTTTGTTTAAAGCAAAACATCTCTCCGATTCATCCCATTCGGTCAATCG	360
Db	301	TCATAGACCTTTGTTTAAAGCAAAACATCTCTCCGATTCATCCCATTCGGTCAATCG	360
Qy	361	GAAGAGAAATTTTGAATCCATTTTCGACAAATAGACCAAGCTCGAAATCCATGCAATG	420
Db	361	GAAGAGAAATTTTGAATCCATTTTCGACAAATAGACCAAGCTCGAAATCCATGCAATG	416
Qy	421	GAAATGAGAGAGATCCTCATATGAGTTTTCGAAATACATGTAATTCGACTCATTAACATA	480
Db	417	GAAATGAGAGAGATCCTCATATGAGTTTTCGAAATACATGTAATTCGACTCATTAACATA	476
Qy	481	GGTGATGTGAATCAAAATGACCTCATGCSCTATCTCTTGGGTATTAACCAAAATAT	540
Db	477	GGTGATGTGAATGAAATGACCCCTCATGCSCTATCTCTTGGGTATTAACCAAAATAT	536
Qy	541	GAGATGAGCCCTGCTCTGATACCAATTTGATGATCAGAGTGGCACCTAAGAGAGGGGG	600
Db	537	GAGATGAGCCCTGCTCTGATACCAATTTGATGATCAGAGTGGCACCTAAGAGAGGGGG	596
Qy	601	GAGTGAATGAGTGGATGATTAACCTTATAGTTTAAATGAAATGAAATCGTAAATACGAG	660
Db	597	GAGTGAATGAGTGGATGATTAACCTTATAGTTTAAATGAAATGAAATCGTAAATACGAG	656
Qy	661	AAGATTTGGTTTAAATGATTAAGTCTGAGTAGATGAAACCAAAAGTTAAACAGTAGTAA	720
Db	657	AAGATTTGGTTTAAATGATTAAGTCTGAGTAGATGAAACCAAAAGTTAAACAGTAGTAA	716
Qy	721	TAAACAATTCGGGAAAGTAAGAACTCACACATTCGAAGAACATACCAATTTAAAGTGGT	780
Db	717	TAAACAATTCGGGAAAGTAAGAACTCACACATTCGAAGAACATACCAATTTAAAGTGGT	776
Qy	781	CGGTCAAAATGACCTTACATCCACTTGTGAGAGCTTCTTCGAGAGAGCTCCCAACTCCAC	840
Db	777	CGGTCAAAATGACCTTACATCCACTTGTGAGAGCTTCTTCGAGAGAGCTCCCAACTCCAC	836
Qy	841	TAGCAAAATCACCTTTCGAAAGGGAGGACAAATACCTCTCTTACNACCTTTTCAATGGTTC	900
Db	837	TAGCAAAATCACCTTTCGAAAGGGAGGACAAATACCTCTCTTACNACCTTTTCAATGGTTC	896
Qy	901	ATACCTTTTCAAAATTTTCAACGAGAAAGAGAGTGAAATGCAATGCAAGCAATTTGAAACAA	960
Db	897	ATACCTTTTCAAAATTTTCAACGAGAAAGAGAGTGAAATGCAATGCAAGCAATTTGAAACAA	956

Qy	961	GACTTGCTAAAGACTTTTGCTAAAGCTTTTCTCAATCTATCTGCTTCTCAAAAGTTGTA	1020
Db	957	GACTTGCTAAAGACTTTTGCTAAAGCTTTTCTCAATCTATCTGCTTCTCAAAAGTTGTA	1016
Qy	1021	TTCTCTGCTGAGAATTCGAGGGTATTATAGACCCCAAGAGATTTAAATTTGGGCTCCA	1080
Db	1017	TTCTCTGCTGAGAATTCGAGGGTATTATAGACCCCAAGAGATTTAAATTTGGGCTCCA	1076
Qy	1081	AAATTCGAAATGCTCTTGGGTTCCCGAGTTGCGGTGCCACCGCTGTCACTGTGTTGACA	1140
Db	1077	AAATTCGAAATGCTCTTGGGTTCCCGAGTTGCGGTGCCACCGCTGTCACTGTGTTGACA	1136
Qy	1141	CTGACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGTGTTGGCGGTGCCACCGC	1200
Db	1137	CTGACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGTGTTGGCGGTGCCACCGC	1196
Qy	1201	CTAGACTTTTTCAGCTCACTGGTTGGATTTCCAAACTTTGACCCCAACAGTCCGAACTCGG	1260
Db	1197	CTAGACTTTTTCAGCTCACTGGTTGGATTTCCAAACTTTGACCCCAACAGTCCGAACTCGG	1256
Qy	1261	GTCCAAATGACCCGTAACCGGATTTATAGGATTAACCTTAATCTTAACCTTAATATATG	1320
Db	1257	GTCCAAATGACCCGTAACCGGATTTATAGGATTAACCTTAATCTTAACCTTAATATATG	1316
Qy	1321	CAAACTAGCGAACTGAAATATAGTCTTAAGCAAGTTTAAACCGGCAACGTCGAGTCT	1380
Db	1317	CAAACTAGCGAACTGAAATATAGTCTTAAGCAAGTTTAAACCGGCAACGTCGAGTCT	1376
Qy	1381	TCCTCCCGGCACTTTTCGGCAGACTTCTGATATACCTTTGGATTTCTCTAGCGGACTCC	1440
Db	1377	TCCTCCCGGCACTTTTCGGCAGACTTCTGATATACCTTTGGATTTCTCTAGCGGACTCC	1436
Qy	1441	TAGTAGGTCGCGATCTTTGTCGCGAGTTTACGAGTAGCCGAACTTCTCGGTGATCTCC	1500
Db	1437	TAGTAGGTCGCGATCTTTGTCGCGAGTTTACGAGTAGCCGAACTTCTCGGTGATCTCC	1496
Qy	1501	GCAACCGCGCATGATCTTTTCGGCAGACTTTTCGAAACTTTTCGAACTTTCGAACTTCT	1560
Db	1497	GCAACCGCGCATGATCTTTTCGGCAGACTTTTCGAAACTTTTCGAACTTTCGAACTTCT	1556
Qy	1561	TCTCGGTTGGTTCGACAGACTCTTAAACGAACTTTCGACTCTTGAATTCGAACTCGAA	1620
Db	1557	TCTCGGTTGGTTCGACAGACTCTTAAACGAACTTTCGAACTTTCGAACTTCGAACTCGAA	1616
Qy	1621	CTTGACTCGGTAGGCTTGTATATTTTCAGGCTATCATAGTAACTTCTACATACATTA	1680
Db	1617	CTTGACTCGGTAGGCTTGTATATTTTCAGGCTATCATAGTAACTTCTACATACATTA	1676
Qy	1681	ACTCAATAATATGATTAAGTAAATTAACCCATCAATGATTTTCATCATCAAAATTCGAC	1740
Db	1677	ACTCAATAATATGATTAAGTAAATTAACCCATCAATGATTTTCATCATCAAAATTCGAC	1736
Qy	1741	ATTCACAAACATCCGTACTCAATAACCCATCAGGCTATAGTACGTGACTATCTACTGT	1800
Db	1737	ATTCACAAACATCCGTACTCAATAACCCATCAGGCTATAGTACGTGACTATCTACTGT	1796
Qy	1801	GATCGGTACGTGAAGTTAGCGAGTCAATCCAGTCTGCTGCTACTTATTTGGCCGACAG	1860
Db	1797	GATCGGTACGTGAAGTTAGCGAGTCAATCCAGTCTGCTGCTACTTATTTGGCCGACAG	1856
Qy	1861	TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCTTACCGCTCTCTTTTTTAT	1920
Db	1857	TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCTTACCGCTCTCTTTTTTAT	1916
Qy	1921	ACTTTTGAAGAAATTCAAATCAAAACAGATTAACAAATTAACAGGTGAGACATGTGACAT	1980
Db	1917	ACTTTTGAAGAAATTCAAATCAAAACAGATTAACAAATTAACAGGTGAGACATGTGACAT	1976
Qy	1981	GCTAGTCTCTGAAAGCAATTAATTCGCGCATCCAGACGTCGTGCTGCTTCTATCCAC	2040
Db	1977	GCTAGTCTCTGAAAGCAATTAATTCGCGCATCCAGACGTCGTGCTGCTTCTATCCAC	2036



QY	2041	TTTTTCTACAT-ACCATGTCGATGGCTTTGTTGATGACAGACCAACAAAGCTTGCCCT	2099
DB	2037	TTTTTCTACATAACCATGTCGATGGCTTTGTTGATGACAGACCAACAAAGCTTGCCCT	2096
QY	2100	TTGTTGTGCGCTTAACAGAGAGAGAGAGAGACGATAGCTCTCATTCACCATG	2159
DB	2097	TTGTTGTGCGCTTAAC-AGAGAGAGAGAGAGACGATAGCTCTCATTCACCATG	2152
QY	2160	G 2160	
DB	2153	G 2153	
RESULT 2			
AAV28690			
ID	AAV28690 standard; cDNA; 809 BP.		
AC	AAV28690;		
XX	29-JUL-1998 (first entry)		
DT	Ripening banana pulp cDNA clone U-D86 SEQ ID NO:48.		
DE	Banana; ripening; pulp; Musa acuminata cv. Grand Nain; fruit;		
KW	Genetic control; tissue senescence; ss.		
KW	Musa acuminata.		
OS	WO9811228-A2.		
PN	19-MAR-1998.		
XX	08-SEP-1997; 97WO-GB002424.		
XX	10-SEP-1996; 96GB-00018962.		
PR	25-APR-1997; 97GB-00008366.		
XX	(ZENE ) ZENECA LTD.		
PA	Seymour GB, Bird CR, Medina-Suarez RDJ;		
XX	WPI; 1998-207389/18.		
XX	Modulation of ripening or tissue senescence in bananas - comprises use of		
PT	DNA isolated from ripening banana pulp to produce genetically modified		
PT	fruit.		
XX	Claim 1; Page 48-49; 72pp; English.		
PS	The present sequence represents a cDNA clone isolated from ripening		
XX	banana pulp. 57 clones were isolated and are given in AAV28643 to		
CC	AAV28699. The cDNA clone sequences can be used in a method of modulating		
CC	ripening or tissue senescence process in plants of the genus Musa. The		
CC	method comprises: (a) inserting into the plant material at least 1 of the		
CC	57 sequences (as above); (b) regenerating the plant material, and (c)		
CC	selecting from the transformed regenerants, plants with modulated		
CC	ripening or tissue senescence characteristics. Also described in the		
CC	present invention are: (1) plants, their progeny, seed and material		
CC	obtained from the plants, produced by the above method; (2) a vector		
CC	functional in plants comprising a promoter region which is operably in		
CC	plant cells, a polynucleotide sequence as defined above, and a		
CC	transcription termination sequence; and (3) a method of controlling plant		
CC	pathogens comprising the application of anti-pathogenic agent to plants		
CC	of (1)		
XX	Sequence 809 BP; 191 A; 233 C; 197 G; 180 T; 0 U; 8 Other;		
QY	Query Match 3.4%; Score 73.4; DB 2; Length 809;		
XX	Best Local Similarity 98.7%; Pred. No. 3.7e-10;		
XX	Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	2086	CCACAGCTTGCTTGGTTGTCCTTAACAGAGAGAGAGAGACGATAGCTT	2145







Query Match 2.0%; Score 43.4; DB 2; Length 742;  
Best Local Similarity 94.9%; Pred. No. 0.14;  
Matches 56; Conservative 0; Mismatches 1; Indels 2; Gaps 1;  
QY 2102 GGTGTGGCTTACAGAGACAGACAGACACCGATAGCTCTCTCATTCACCATGG 2160

XX  
SQ Sequence 1186 BP; 277 A; 325 C; 295 G; 289 T; 0 U; 0 Other;  
Query Match 2.0%; Score 42.4; DB 2; Length 1186;  
Best Local Similarity 91.9%; Pred.No. 0.35;  
Matches 57; Conservative 0; Mismatches 1; Indels 4; Gaps 1



one or more data sets. The genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplification carries a fluorescent label or radionuclide. Optionally, the labels of the amplicates are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplicates or fragments of the amplicates are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 7001 BP; 2309 A; 40 C; 1249 G; 3403 T; 0 U; 0 Other;

Query Match 1.9%; Score 41.6; DB 6; Length 7001;  
Best Local Similarity 48.0%; Pred. No. 1.5;  
Matches 119; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 229 AAATATCAAAAGCGGATCAAAATGGAATGTAATGTAATTCATTTTGTCTAATGCAC 288  
Db 667 AAAATTCAAAATAAATACTTATCCCTAAATACAAAATAATAAATTCATTTCAAATATC 608  
QY 289 AAAACAGGGGATTCATAGCCTTTGTTTAAAGCAAAAACATTTCTCCGATTCATCCC 348  
Db 607 ATTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 548  
QY 349 ATTGCGTCATCGGAGAGAGAAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGA 408  
Db 547 ATCAATAAATAACATTTTAAATATTTAATATAGATATATACTTAACATAAAAACTT 488  
QY 409 AATCCATGATGGAATGAGAGAGATCCTCATATGAGTTTTCGAATACATGTAATTCGAC 468  
Db 487 AAAACACAAACCCCTAATAAACCAATCCTATAATAATTTTATTACACATATATCTTAT 428  
QY 469 TCATTAAA 476  
Db 427 AAAATAAA 420

RESULT 14

ADA20394/c  
ID ADA20394 standard; DNA; 7001 BP.

AC ADA20394;

DT 20-NOV-2003 (first entry)

DE Prostate tumour related genomic DNA sample #30.

XX cytostatic; gene therapy; genetic marker; epigenetic parameter;  
XX classification; differentiation; diagnosis; prostate tumour;  
XX prostate cancer; cytosine methylation; uracil;  
XX single nucleotide polymorphism; SNP; prostate carcinoma, ss.

OS Homo sapiens.

XX WO2002103042-A2.

PN 27-DEC-2002.

XX 14-JUN-2002; 2002WO-EP006605.

XX 14-JUN-2001; 2001DE-01028508.

XX (EPG-) EPIGENOMICS AG.

PA Distler J, Model F, Adorjan P;

PI

XX

WPI; 2003-167536/16.

Determining genetic and/or epigenetic parameters, useful for the classification, differentiation and/or diagnosis of prostate tumors or a predilection to prostate cancer, comprises analyzing cytosine methylation.

Claim 28; Page 262-266; 376pp; English.

The invention relates to a method of determining genetic and/or epigenetic parameters for the classification, differentiation and/or diagnosis of prostate tumors or the predilection to prostate cancer, by analysing cytosine methylation in a sample of genomic DNA. The method comprises chemically treating unmethylated cytosine bases at the 5-position to uracil or another base, which is dissimilar to cytosine in terms of hybridization behaviour; followed by amplifying at least one fragment of the chemically pre-treated genomic DNA using sets of primer oligonucleotides and a polymerase. The oligomers or probes derived from them are useful for detecting the methylation state of all CpG dinucleotides and/or single nucleotide polymorphisms (SNPs) in a chemically pre-treated genomic DNA. They are all useful for treating prostate carcinoma. This sequence represents a fragment of genomic DNA used in the method of the invention.

Sequence 7001 BP; 2309 A; 40 C; 1249 G; 3403 T; 0 U; 0 Other;

Query Match 1.9%; Score 41.6; DB 7; Length 7001;  
Best Local Similarity 48.0%; Pred. No. 1.5;  
Matches 119; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 229 AAATATCAAAAGCGGATCAAAATGGAATGTAATGTAATTCATTTTGTCTAATGCAC 288  
Db 667 AAAATTCAAAATAAATACTTATCCCTAAATACAAAATAATAAATTCATTTCAAATATC 608  
QY 289 AAAACAGGGGATTCATAGCCTTTGTTTAAAGCAAAAACATTTCTCCGATTCATCCC 348  
Db 607 ATTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 548  
QY 349 ATTGCGTCATCGGAGAGAGAAAATTTTGAATCCATTTTCGACATAGACCAAGCTCGA 408  
Db 547 ATCAATAAATAACATTTTAAATATTTAATATAGATATATACTTAACATAAAAACTT 488  
QY 409 AATCCATGATGGAATGAGAGAGATCCTCATATGAGTTTTCGAATACATGTAATTCGAC 468  
Db 487 AAAACACAAACCCCTAATAAACCAATCCTATAATAATTTTATTACACATATATCTTAT 428  
QY 469 TCATTAAA 476  
Db 427 AAAATAAA 420

RESULT 15

ADA84201/c

ID ADA84201 standard; DNA; 7001 BP.

XX ADA84201;

DT 20-NOV-2003 (first entry)

DE Human renal/prostate carcinoma associated DNA SEQ ID NO:59.

XX ds; renal cancer; prostate cancer; cytosine methylation;

XX single nucleotide polymorphism; histological; cytological.

OS Homo sapiens.

XX WO2002103041-A2.

PN 27-DEC-2002.

XX 14-JUN-2002; 2002WO-EP006603.

XX





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:51:07 ; Search time 152.641 Seconds  
(without alignments)  
7853.010 Million cell updates/sec

Title: US-09-892-635A-45

Perfect score: 2160

Sequence: 1 ggatcccaacttttaggaat.....agctctctcattcaccatgg 2160

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	62.8	2.9	2417	US-09-254-776B-3	Sequence 3, Appli
C 2	52.6	2.4	515	US-08-282-581-5	Sequence 5, Appli
C 3	52.6	2.4	515	US-08-550-544-5	Sequence 5, Appli
C 4	48.4	2.2	7218	US-08-232-463-14	Sequence 14, Appl
C 5	44.8	2.1	7218	US-08-232-463-14	Sequence 14, Appl
C 6	42.8	2.0	832	US-09-621-976-2813	Sequence 2813, Ap
C 7	41.2	1.9	2881	US-08-956-171B-254	Sequence 254, App
C 8	37.6	1.7	5335	US-10-204-708-17	Sequence 17, Appl
C 9	37	1.7	6317	US-10-204-708-11	Sequence 11, Appl
C 10	36.6	1.7	399	US-09-621-976-8976	Sequence 8976, Ap
C 11	36.6	1.7	2871	US-09-023-655-768	Sequence 768, App
C 12	36.2	1.7	1797	US-09-134-000C-1552	Sequence 1552, Ap
C 13	36.2	1.7	6113	US-10-204-708-13	Sequence 13, Appl
C 14	35.8	1.7	6124	US-08-213-419B-3	Sequence 3, Appli
C 15	35.8	1.7	21234	US-09-810-671-3	Sequence 3, Appli
C 16	35.8	1.7	21234	US-10-109-854-3	Sequence 3, Appli
C 17	35.6	1.6	1704	US-09-134-001C-925	Sequence 925, App
C 18	35.6	1.6	168575	US-09-426-290-1	Sequence 1, Appli
C 19	35.4	1.6	4185	US-09-417-485D-7	Sequence 7, Appli
C 20	35.4	1.6	10640	US-09-417-485D-5	Sequence 5, Appli
C 21	35.2	1.6	11050	US-10-204-708-86	Sequence 86, Appl
C 22	35	1.6	6656	US-10-204-708-76	Sequence 76, Appl
C 23	35	1.6	10467	US-10-204-708-2	Sequence 2, Appli
C 24	35	1.6	1664976	US-08-916-421B-1	Sequence 1, Appli
C 25	34.8	1.6	3440	US-08-471-791-27	Sequence 27, Appl
C 26	34.8	1.6	3440	PCT-US91-01746-27	Sequence 27, Appl
C 27	34.8	1.6	5562	US-10-204-708-64	Sequence 64, Appl

C 28 34.6 1.6 627 4 US-09-495-050A-74 Sequence 74, Appli  
C 29 34.6 1.6 6020 4 US-10-204-708-8 Sequence 8, Appli  
C 30 34.4 1.6 832 4 US-09-621-976-2813 Sequence 2813, Ap  
C 31 34.4 1.6 887 3 US-09-602-848-3 Sequence 3, Appli  
C 32 34.4 1.6 99500 4 US-09-798-096-10 Sequence 10, Appl  
C 33 34.2 1.6 337 1 US-09-985-799-79 Sequence 79, Appl  
C 34 34.2 1.6 337 1 US-08-594-031-79 Sequence 79, Appl  
C 35 34.2 1.6 2427 4 US-09-134-001C-269 Sequence 268, App  
C 36 34.2 1.6 4673 1 US-07-638-431-1 Sequence 1, Appli  
C 37 34.2 1.6 4673 5 PCT-US92-00018-1 Sequence 1, Appli  
C 38 34.2 1.6 5152 4 US-10-204-708-74 Sequence 74, Appl  
C 39 34.2 1.6 580073 4 US-08-545-528B-1 Sequence 1, Appli  
C 40 34.2 1.6 640681 4 US-09-790-988-1 Sequence 1, Appli  
C 41 34 1.6 505 4 US-09-621-976-15639 Sequence 15639, A  
C 42 34 1.6 576 4 US-09-621-976-16902 Sequence 16902, A  
C 43 34 1.6 1991 3 US-08-235-836C-69 Sequence 69, Appl  
C 44 34 1.6 1991 3 US-08-235-836C-77 Sequence 77, Appl  
C 45 34 1.6 5521 3 US-08-975-762-48 Sequence 48, Appl

## ALIGNMENTS

RESULT 1  
US-09-254-776B-3/c  
; Sequence 3, Application US/09254776B  
; Patent No. 6559359  
; GENERAL INFORMATION:  
; APPLICANT: Laten, Howard  
; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF  
; FILE REFERENCE: 27013/33479A  
; CURRENT APPLICATION NUMBER: US/09/254,776B  
; CURRENT FILING DATE: 1999-03-09  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 2417  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: SIRE-1 cdna clone  
US-09-254-776B-3

Query Match 2.9%; Score 62.8; DB 4; Length 2417;  
Best Local Similarity 50.6%; Pred. No. 9.7e-09;  
Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;

QY 4 TCCCACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAGTTAGAAAATCT 63  
DB 786 TCAATCTTTTAGGAAGGATCTGAGGATCTTCTCACCAGCTTTTCATCAGTCATCTT 727  
QY 64 TTACCAAGAGCTTTGAGTCCATTGATGACATC-CGTGAAACGGTGACATGTCCTCGATG 122  
DB 726 TCTCCCAAGGAGTGAAGCATTCGCAATTCACGAATGTTTCATGTGAAGTCAATGA 667  
QY 123 GACTCAGTGGTTTCATTCGGAAAAGTTCGAAAGAGTGCATGAATATTTGATTTGGAT 182  
DB 666 CACTCTTCCTCTTCATCTTCAGATTTTCGAATTTTGTGGCCAAATAGTTGCAATCTGGAC 607  
QY 183 TCTTTCACTCGGTTGGTGGCTTCATGAGTGCCTCAAGAGTCTCAATATCAAAAGCC 242  
DB 606 ATCTTCACCTTTGGAGGTTCCCTTCATGAGTGGTTTTCAGGATCTCCCATGCTCTTGGCC 547  
QY 243 GAATCACAATGAATGTGATTGAATTCATTTTGTCTAATGACACAAACAGGCAATTC 302  
DB 546 ACTGTGCAATGTGTCATCAGTCTGAGATATTTCTGTCACTCCATTCCTGATGAGCATTC 487  
QY 303 ATAGCTTTGTGTTTAAAGCAAAAACATTTCTCCGATTTTCATCCCATTC 352  
DB 486 AAAGCTTTGGAGTTTCCCAAGTGCCAATTCGCTCTTCTTCTTTAGTCCAGTC 437

RESULT 2  
US-08-282-581-5  
; Sequence 5, Application US/08282581  
; Patent No. 5670349  
; GENERAL INFORMATION:  
; APPLICANT: Cramer, Carole L.  
; APPLICANT: Weissenborn, Deborah L.  
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND  
; TITLE OF INVENTION: POST-HARVEST PRODUCTION OF GENE PRODUCTS  
; TITLE OF INVENTION: IN PLANTS AND PLANT CELL CULTURES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/550,544  
; FILING DATE: US/08/550,544  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/100,816  
; FILING DATE: 02-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18872  
; REFERENCE/DOCKET NUMBER: 7956-004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-869-8864/9741  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 515 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-08-550-544-5  
Query Match 2.4%; Score 52.6; DB 1; Length 515;  
Best Local Similarity 51.5%; Pred. No. 5.1e-06;  
Matches 121; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
QY 135 TTCATTCGAAAGTTCGAAAGAGTGCATAGATATTTGGATTCTTCACTCGG 194  
Db 122 TTCATTTTGAAGTTCATATCGGTGAGCATGTCATCTTGGATTCTTGACTTGT 181  
QY 195 TTGTGCTTCATGAGTCACTCAAGAGTCTCCAAATATCAAAAGCCGATCAAAATT 254  
Db 182 TCAGTTTCTCATGTGTAGTCAACAGCAATCCAGATTCTTTAGCAGACTCAGGCT 241  
QY 255 GAAATGCTGATTAATTCATTTTCTTAATGCACAAACAGGCAATCATAGCCTTTGTG 314  
Db 242 GACACTCTATGTACTCATCAGGTCTTATCCACAGACCATAGAGTTTACCTTTGAAG 301  
QY 315 TTTAAAGCAAAACAAATTCTTCCTCGATTCCCATTCGCTCATCGGGAAGAAA 369  
Db 302 CCCTTTCTATCTTTTCTCTGTCAGCATCATATTTCTGCTGGGCTTTGGAA 356  
RESULT 4  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA

US-08-282-581-5  
; Sequence 5, Application US/08282581  
; Patent No. 5670349  
; GENERAL INFORMATION:  
; APPLICANT: Cramer, Carole L.  
; APPLICANT: Weissenborn, Deborah L.  
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM AND  
; TITLE OF INVENTION: POST-HARVEST PRODUCTION OF GENE PRODUCTS  
; TITLE OF INVENTION: IN PLANTS AND PLANT CELL CULTURES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/282,581  
; FILING DATE: 29-JUL-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18872  
; REFERENCE/DOCKET NUMBER: 7956-009  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-869-8864/9741  
; TELEFAX: 212-869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 515 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; US-08-282-581-5  
Query Match 2.4%; Score 52.6; DB 1; Length 515;  
Best Local Similarity 51.5%; Pred. No. 5.1e-06;  
Matches 121; Conservative 0; Mismatches 114; Indels 0; Gaps 0;  
QY 135 TTCATTCGAAAGTTCGAAAGAGTGCATAGATATTTGGATTCTTCACTCGG 194  
Db 122 TTCATTTTGAAGTTCATATCGGTGAGCATGTCATCTTGGATTCTTGACTTGT 181  
QY 195 TTGTGCTTCATGAGTCACTCAAGAGTCTCCAAATATCAAAAGCCGATCAAAATT 254  
Db 182 TCAGTTTCTCATGTGTAGTCAACAGCAATCCAGATTCTTTAGCAGACTCAGGCT 241  
QY 255 GAAATGCTGATTAATTCATTTTCTTAATGCACAAACAGGCAATCATAGCCTTTGTG 314  
Db 242 GACACTCTATGTACTCATCAGGTCTTATCCACAGACCATAGAGTTTACCTTTGAAG 301  
QY 315 TTTAAAGCAAAACAAATTCTTCCTCGATTCCCATTCGCTCATCGGGAAGAAA 369  
Db 302 CCCTTTCTATCTTTTCTCTGTCAGCATCATATTTCTGCTGGGCTTTGGAA 356  
RESULT 3  
US-08-550-544-5  
; Sequence 5, Application US/08550544  
; Patent No. 5689056  
; GENERAL INFORMATION:  
; APPLICANT: Cramer, Carole L.  
; APPLICANT: Weissenborn, Deborah L.  
; TITLE OF INVENTION: HMG2 PROMOTER EXPRESSION SYSTEM







Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 768:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2871 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTTUT01  
CLONE: 602872  
US-09-023-655-768  
Query Match 1.7% Score 36.6; DB 4; Length 2871;  
Best Local Similarity 45.5%; Pred. No. 1.2;  
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;  
Qy 514 ATCTCTTGGGTATTAACCAATATGAGAGTGAGCCTTCTCTGATACCAATTGTTAG 573  
Db 717 ATATTACAGGGTCAAGAAACAAAGCTGTCTCCAGTCATGTTTGGACAATAACGTTTGG 658  
Qy 574 GATCAGAGTGGCACTAAGAGAGGGGGAGTGAATTAGTCAGTCGATTAAACCTTAA 633  
Db 657 GGGTCAGAGCGGAAGAGGAGGAGAAAGAGAAAGAGAGAGGAGAAATACACTACTT 598  
Qy 634 GTTTAAAAAATGAATTCGTAATACGAGAGAGATTTTCGTTTTAATAGTAACTTGAGTAG 693  
Db 597 CTGGAAAAACACATTTGGCTTAACTCCAAAATAAAGGCTTTTCGCGAGAAATGAAAGCCT 538  
Qy 694 AAAACCAAGTTAAGTAGTAGTGAATAACAAATTCGGGAAAGTAAGAACTCACACATT 753  
Db 537 ATAATCAGGATTTAGTGTGCAATAAACAACACAGCTGCACACAGACAATATCTACATCAT 478  
Qy 754 CAAGGAACATACCAATTTAAAGTGGTTCCGTCACAAAT 790  
Db 477 TAATANCCTTCNATTATAGATCTATCGTCACAAAT 441  
RESULT 12  
US-09-134-000C-1552  
Sequence 1552, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1552  
LENGTH: 1797  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1552  
Query Match 1.7% Score 36.2; DB 4; Length 1797;  
Best Local Similarity 49.2%; Pred. No. 1.2;  
Matches 95; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
Qy 926 AAGAAGGAGGTGAACATCAAGCAATTGAAAAACAAGACTTGCTAAAGACTTTGCTAAGGC 985  
Db 761 AAGAAGGAGACCAAGAAAGCTAATGAATTATGGAATGGTTCGCTAGCGAGTCTCTAAAA 820  
Qy 986 TTTTCTTCTCAATCTATTGCTTCTCAAAAGTTGTTCTCTGCTGAGAAATGAGGGTAT 1045  
Db 821 AATTGATAAAATCTATTCCATGTTAGAAAGTTGAATTTGATTTTATTAATGGCGAAGCCT 880  
Qy 1046 TTATAGACCCCAAGAGGATTTAAATTTGGGTCCTCAAAATTTGGAATGCTTTGGGTTCCCG 1105  
Db 881 TTTATAATGATAAATGATGAATTTGTTACTTTTATTAGAGAAAAACAATTTATTACAG 940  
Qy 1106 AGGTTGCGGTGC 1118  
Db 941 AAAATCAAGGTGC 953

Sequence 1552, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1552  
LENGTH: 1797  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-1552  
Query Match 1.7% Score 36.2; DB 4; Length 1797;  
Best Local Similarity 49.2%; Pred. No. 1.2;  
Matches 95; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
Qy 926 AAGAAGGAGGTGAACATCAAGCAATTGAAAAACAAGACTTGCTAAAGACTTTGCTAAGGC 985  
Db 761 AAGAAGGAGACCAAGAAAGCTAATGAATTATGGAATGGTTCGCTAGCGAGTCTCTAAAA 820  
Qy 986 TTTTCTTCTCAATCTATTGCTTCTCAAAAGTTGTTCTCTGCTGAGAAATGAGGGTAT 1045  
Db 821 AATTGATAAAATCTATTCCATGTTAGAAAGTTGAATTTGATTTTATTAATGGCGAAGCCT 880  
Qy 1046 TTATAGACCCCAAGAGGATTTAAATTTGGGTCCTCAAAATTTGGAATGCTTTGGGTTCCCG 1105  
Db 881 TTTATAATGATAAATGATGAATTTGTTACTTTTATTAGAGAAAAACAATTTATTACAG 940  
Qy 1106 AGGTTGCGGTGC 1118  
Db 941 AAAATCAAGGTGC 953  
RESULT 13  
US-10-204-708-13  
Sequence 13, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 13  
LENGTH: 6113  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-13  
Query Match 1.7% Score 36.2; DB 4; Length 6113;  
RESULT 12  
US-09-134-000C-1552  
Sequence 1552, Application US/09134000C  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 768:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2871 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTTUT01  
CLONE: 602872  
US-09-023-655-768  
Query Match 1.7% Score 36.6; DB 4; Length 2871;  
Best Local Similarity 45.5%; Pred. No. 1.2;  
Matches 126; Conservative 0; Mismatches 151; Indels 0; Gaps 0;  
Qy 514 ATCTCTTGGGTATTAACCAATATGAGAGTGAGCCTTCTCTGATACCAATTGTTAG 573  
Db 717 ATATTACAGGGTCAAGAAACAAAGCTGTCTCCAGTCATGTTTGGACAATAACGTTTGG 658  
Qy 574 GATCAGAGTGGCACTAAGAGAGGGGGAGTGAATTAGTCAGTCGATTAAACCTTAA 633  
Db 657 GGGTCAGAGCGGAAGAGGAGGAGAAAGAGAAAGAGAGAGGAGAAATACACTACTT 598  
Qy 634 GTTTAAAAAATGAATTCGTAATACGAGAGAGATTTTCGTTTTAATAGTAACTTGAGTAG 693  
Db 597 CTGGAAAAACACATTTGGCTTAACTCCAAAATAAAGGCTTTTCGCGAGAAATGAAAGCCT 538  
Qy 694 AAAACCAAGTTAAGTAGTAGTGAATAACAAATTCGGGAAAGTAAGAACTCACACATT 753  
Db 537 ATAATCAGGATTTAGTGTGCAATAAACAACACAGCTGCACACAGACAATATCTACATCAT 478  
Qy 754 CAAGGAACATACCAATTTAAAGTGGTTCCGTCACAAAT 790  
Db 477 TAATANCCTTCNATTATAGATCTATCGTCACAAAT 441  
RESULT 12  
US-09-134-000C-1552

	Best Local Similarity	49.2%;	Pred. No. 2.6;						
	Matches	95;	Conservative	0;	Mismatches	98;	Indels	0;	Gaps
QY	595	GGGGGGAGTGAAT	TAGTGCAGTGGAT	TAAAACTT	TAACTTT	AAAAATGAAT	TGTTAAA	654	
Db	906	GGTGTTGGTTAT	TAAATATAT	TATCTGTT	TAAATTT	TATAGAATT	TATATTAATTT	965	
QY	655	TACGAGAAGATTT	CGTTTTT	TAATAGTACT	TGAGTAGAT	GAAACCAAAAGTT	TACAGTAG	714	
Db	966	GTAAATTTTAT	CTAATTT	AAAAATAGAA	TTTAAAAAAAT	TATTAATAGTAAT	TAGAAA	1025	
QY	715	TGTAATAACAAT	TTTCGGGAAAGT	AGAACTC	ACAACTC	ACAGAAACAT	TACCAATTTAAA	774	
Db	1026	ATTAAAGAAAAAA	ATGAGATAT	TATTAATATAT	TTTATT	TAGAAGAGTT	AAAAATTTAAAAGA	1085	
QY	775	GTGGTTCCGTC	CAA					787	
Db	1086	TTGATTATAT	TAA					1098	

RESULT 14  
 US-08-213-419B-3  
 ; Sequence 3, Application US/08213419B  
 ; Patent No. 6333406  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Inselburg, J. et al.  
 ; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; FILE REFERENCE: J11-002CNCIP  
 ; CURRENT APPLICATION NUMBER: US/08/213,419B  
 ; CURRENT FILING DATE: 1994-03-14  
 ; PRIOR APPLICATION NUMBER: US 07/870,506  
 ; PRIOR FILING DATE: 1992-04-17  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 6124  
 ; TYPE: DNA  
 ; ORGANISM: Plasmodium falciparum  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (2407)..(2439)  
 ; NAME/KEY: CDS  
 ; LOCATION: (2598)..(3404)  
 ; NAME/KEY: CDS  
 ; LOCATION: (3580)..(3720)  
 ; NAME/KEY: CDS  
 ; LOCATION: (3850)..(5835)  
 US-08-213-419B-3

[illegible]

RESULT 15  
US-09-810-671-3/c  
; Sequence 3, Application US/09810671  
; Patent No. 6455291  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al

```

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000758
; CURRENT APPLICATION NUMBER: US/09/810,671
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 21234
; TYPE: DNA
; ORGANISM: Human
US-09-810-671-3

```

Query Match	1.7%;	Score 35.8;	DB 4;	Length 21234;
Best Local Similarity	48.7%;	Pred. No. 7.4;		
Matches	97;	Conservative 0;	Mismatches 102;	Indels 0; Gaps 0;
QY	578	AGAGTGGCACTAAGACAGAGGGGGGAGTGAATTAGTCAGTGGATTAACAACTTATAAGTTT	637	
Db	6518	AGGTTAACATTTGGAAATATATGGGAATCTTTGAACTATTTTACATTTTCTGTAAGT	6459	
QY	638	AAAAATGAAATTCGTAAATACGAAAGATTTCGTTTTAATAGTAACCTTGAGTAGATGAAAA	697	
Db	6458	CTGAAAATATTTCAAATATACTTTAAAGTTCAATTTACATTCACATATTTGTCGAGAAA	6399	
QY	698	CAAAAAGTTTAAACAGTAGTGTAAATAACAATTTTCGGGAAGTAGAAGTACACATTCAG	757	
Db	6398	GCAAAATGTGGGATTAATTTAGGATAGTATTTATGGAAAGTGAATGACANTAAATTCAA	6339	
QY	758	GAACATACCAATTTTAAAGT	776	
Db	6338	TAATAAAACATTTCTAAAGT	6320	

Search completed: September 18, 2004, 14:23:51  
Job time : 155.641 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 14:07:43 ; Search time 968.897 Seconds  
(without alignments)  
11252.488 Million cell updates/sec

Title: US-09-892-635A-45  
Perfect score: 2160  
Sequence: 1 ggatcccaacttttaggaat.....agctctcattcaccatgg 2160

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2158.6	99.9	2160	13	US-09-892-635A-45
2	2139	99.0	2156	13	Sequence 45, Appl
3	2110	97.7	4924	13	Sequence 44, Appl
4	122.2	5.7	78333	16	Sequence 27, Appl
5	81.4	3.8	2760	13	Sequence 3, Appl
6	81.2	3.8	58955	16	Sequence 108338
7	73.4	3.4	809	9	Sequence 2, Appl
8	62.8	2.9	2417	15	Sequence 48, Appl
9	62.8	2.9	2417	15	Sequence 3, Appl
10	62.8	2.9	2801	13	Sequence 3, Appl
11	62.8	2.9	9399	16	Sequence 108650
12	61.2	2.8	5371	17	Sequence 93, Appl
13	59.6	2.8	9072	16	Sequence 9498, Ap
14	59.4	2.8	5016	17	Sequence 87, Appl
					Sequence 10782, A

C 15	59.4	2.8	5304	17	US-10-437-963-44132	Sequence 44132, A
C 16	57.8	2.7	3843	17	US-10-437-963-10117	Sequence 10117, A
C 17	57.8	2.7	4326	17	US-10-437-963-9179	Sequence 9179, Ap
C 18	57.8	2.7	4428	17	US-10-437-963-44137	Sequence 44137, A
C 19	57.8	2.7	4840	17	US-10-437-963-9223	Sequence 9223, Ap
C 20	57.8	2.7	10369	17	US-10-437-963-10605	Sequence 10605, A
C 21	57.6	2.7	4503	17	US-10-437-963-8782	Sequence 8782, Ap
C 22	56.6	2.6	7628	13	US-10-424-599-108640	Sequence 108640, Ap
C 23	56.2	2.6	4002	17	US-10-437-963-9586	Sequence 9586, Ap
C 24	56.2	2.6	4428	17	US-10-437-963-9621	Sequence 9621, Ap
C 25	56	2.6	3127	17	US-10-437-963-44134	Sequence 44134, A
C 26	55.6	2.6	9114	17	US-10-437-963-10728	Sequence 10728, A
C 27	55.6	2.6	2160	17	US-10-437-963-101688	Sequence 101688, A
C 28	54.8	2.5	846	17	US-10-437-963-96678	Sequence 96678, A
C 29	54.6	2.5	4884	17	US-10-437-963-9664	Sequence 9664, Ap
C 30	54.4	2.5	5370	17	US-10-437-963-9697	Sequence 9697, Ap
C 31	53.8	2.5	1731	17	US-10-437-963-9134	Sequence 9134, Ap
C 32	53.4	2.5	4382	17	US-10-437-963-44138	Sequence 44138, A
C 33	53.4	2.5	4716	17	US-10-437-963-10750	Sequence 10750, A
C 34	53.2	2.5	13744	17	US-10-437-963-11092	Sequence 11092, A
C 35	53	2.5	4125	17	US-10-437-963-9312	Sequence 9312, Ap
C 36	53	2.5	4620	17	US-10-437-963-9312	Sequence 10128, A
C 37	53	2.5	6036	17	US-10-437-963-44135	Sequence 44135, A
C 38	52.6	2.4	515	9	US-09-902-653-5	Sequence 5, Appli
C 39	52.6	2.4	515	13	US-10-394-494-5	Sequence 5, Appli
C 40	52.4	2.4	5856	17	US-10-437-963-23904	Sequence 23904, A
C 41	52	2.4	5208	17	US-10-437-963-34842	Sequence 34842, A
C 42	51.8	2.4	3810	17	US-10-437-963-24004	Sequence 24004, A
C 43	51.4	2.4	5092	17	US-10-437-963-10162	Sequence 10162, A
C 44	51.2	2.4	3231	17	US-10-437-963-10084	Sequence 10084, A
C 45	51.2	2.4	4420	17	US-10-437-963-9267	Sequence 9267, Ap

## ALIGNMENTS

## RESULT 1

US-09-892-635A-45  
; Sequence 45, Application US/09892635A  
; Publication No. US20030226175A1  
; GENERAL INFORMATION:  
; APPLICANT: May, Gregory D.  
; APPLICANT: Clendennen, Stephanie K.  
; APPLICANT: Mason, Hugh S.  
; APPLICANT: Lim, Miguel A. Gomez  
; APPLICANT: Arntzen, Charles J.  
; TITLE OF INVENTION: DNA Regulatory Elements Associated with Fruit Development  
; FILE REFERENCE: 031998-007  
; CURRENT APPLICATION NUMBER: US/09/892,635A  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 09/160,351  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: US 60/060,062  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 2160  
; TYPE: DNA  
; ORGANISM: Musa acuminata  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 883  
; OTHER INFORMATION: n = A,T,C or G  
US-09-892-635A-45

Query Match 99.9%; Score 2158.6; DB 13; Length 2160;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCCCACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAGTTAGAAAA 60  
|||||  
DB 1 GGATCCCACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAGTTAGAAAA 60  
|||||

QY 61 TCTTTACAAAGAGCTTTGAGTCCATGATGACATCCGTGAAAACGGTGTATCATGTCTCCGA 120  
Db 61 TCTTTACAAAGAGCTTTGAGTCCATGATGACATCCGTGAAAACGGTGTATCATGTCTCCGA 120  
QY 121 TGGACTCAGTGGTTTTCATTCCGGAAGAGTTGCGAAGAGTGATAGAAATATTGATTTTGG 180  
Db 121 TGGACTCAGTGGTTTTCATTCCGGAAGAGTTGCGAAGAGTGATAGAAATATTGATTTTGG 180  
QY 181 ATTCTTTCACTCGGTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAG 240  
Db 181 ATTCTTTCACTCGGTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAG 240  
QY 241 CCGAATCACAAATTTGAAATGATGATGAAATTCATTTTGTCTAATGACAAAACAGGGCAT 300  
Db 241 CCGAATCACAAATTTGAAATGATGATGAAATTCATTTTGTCTAATGCAAAAACAGGGCAT 300  
QY 301 TCATAGCCTTTGTGTTTAAAGCAAAAACATTTCTCCGATTCATCCCATTCGCTCATCG 360  
Db 301 TCATAGCCTTTGTGTTTAAAGCAAAAACATTTCTCCGATTCATCCCATTCGCTCATCG 360  
QY 361 GAAGAGAAAAATTTTGAATCCATTTTCGACAAATAGACCAAAAGCTCGAAATCCATGCAATG 420  
Db 361 GAAGAGAAAAATTTTGAATCCATTTTCGACAAATAGACCAAAAGCTCGAAATCCATGCAATG 420  
QY 421 GAAATGAGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAACATA 480  
Db 421 GAAATGAGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAACATA 480  
QY 481 GGTGATGTGTAATGAAATGACCCCTCATGCTATCTCTTTGGGTATTAACCAAAATAT 540  
Db 481 GGTGATGTGTAATGAAATGACCCCTCATGCTATCTCTTTGGGTATTAACCAAAATAT 540  
QY 541 GAGAGTGAGCCTTGCTCTGATACCAATTTGTTAGGATCAGAGTGCGCACTAAGAGAGGGGG 600  
Db 541 GAGAGTGAGCCTTGCTCTGATACCAATTTGTTAGGATCAGAGTGCGCACTAAGAGAGGGGG 600  
QY 601 GAGTGAATAGTGAGTGCAGTGAATAAATCTAATAGTTTAAATGTAATTCGTAATACAG 660  
Db 601 GAGTGAATAGTGAGTGCAGTGAATAAATCTAATAGTTTAAATGTAATTCGTAATACAG 660  
QY 661 AAGATTTCGTTTAAATAGTAACCTGAGTAGATGAAACCAAAAGTTAAACAGTAGTGTAA 720  
Db 661 AAGATTTCGTTTAAATAGTAACCTGAGTAGATGAAACCAAAAGTTAAACAGTAGTGTAA 720  
QY 721 TAACAATTTCCGGAAGTAAGAACTCACACATTCGAAGGAACATACCAATTTAAAGTGTT 780  
Db 721 TAACAATTTCCGGAAGTAAGAACTCACACATTCGAAGGAACATACCAATTTAAAGTGTT 780  
QY 781 CGGTCAAAATGACCTTACATCCACTTTGAGAGCTTTCTCGAAGAGGCTCCCAACTTCCAC 840  
Db 781 CGGTCAAAATGACCTTACATCCACTTTGAGAGCTTTCTCGAAGAGGCTCCCAACTTCCAC 840  
QY 841 TAGCAAAATCACCTTTGAAGGGGAGGACAAATACCTCTCTTACNACCTTTTCAATAGTTTC 900  
Db 841 TAGCAAAATCACCTTTGAAGGGGAGGACAAATACCTCTCTTACNACCTTTTCAATAGTTTC 900  
QY 901 ATACTCTTACAAATTTTCAACGAGAAAGAGAGTGGAACATGGAAGCAATTTGAAAAA 960  
Db 901 ATACTCTTACAAATTTTCAACGAGAAAGAGAGTGGAACATGGAAGCAATTTGAAAAA 960  
QY 961 GACTTTGCTAAAGACTTTGCTAGGCTTTTCTCAATCTATTTGCTTCTCAAAAGTTGTA 1020  
Db 961 GACTTTGCTAAAGACTTTGCTAGGCTTTTCTCAATCTATTTGCTTCTCAAAAGTTGTA 1020  
QY 1021 TTCTCTGCTGAGAAATGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCA 1080  
Db 1021 TTCTCTGCTGAGAAATGAGGGGTATTTATAGACCCCAAGAGGATTTAAATTTGGGCTCA 1080  
QY 1081 AATTTTGAATGCTCTTGGGTTCCTCGAGTTGCGGTGCAACCGCTGTGAGTTTGTACA 1140  
Db 1081 AATTTTGAATGCTCTTGGGTTCCTCGAGTTGCGGTGCAACCGCTGTGAGTTTGTACA 1140

RESULT 2

US-09-892-635A-44

; Sequence 44, Application US/09892635A

; Publication No. US20030226175A1

QY 1141 CTGACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGGTGTGTTGGCGGTGCGACCGC 1200  
Db 1141 CTGACAGTGTACTAGCGGTGCCACCGCGGACCTCTCGGGTGTGTTGGCGGTGCGACCGC 1200  
QY 1201 CTAGACTTTTTCAGCTCAGCTGTTGAGTTCAAACTTTGACCCCAACAGTCCGAACTCGG 1260  
Db 1201 CTAGACTTTTTCAGCTCAGCTGTTGAGTTCAAACTTTGACCCCAACAGTCCGAACTCGG 1260  
QY 1261 GTCCAAATGACCGGTAAACCGGATATAGGATTAACCCCTTAATCCTAACCCTAATATATG 1320  
Db 1261 GTCCAAATGACCGGTAAACCGGATATAGGATTAACCCCTTAATCCTAACCCTAATATATG 1320  
QY 1321 CAAACTACGCAACTGAAATATATAGTCCTAAGCAAGTTTAAACCGGCAAAAGCTCGAGTCT 1380  
Db 1321 CAAACTACGCAACTGAAATATATAGTCCTAAGCAAGTTTAAACCGGCAAAAGCTCGAGTCT 1380  
QY 1381 TCTTCCGGCGAATCTTTCGGCAGACTCTGATATACCTTTGGATTTCTCTAGCGGACTCC 1440  
Db 1381 TCTTCCGGCGAATCTTTCGGCAGACTCTGATATACCTTTGGATTTCTCTAGCGGACTCC 1440  
QY 1441 TAGTAGGTCCTCGATCTTGTGGCGAGTTTAGCGAGTAGCCGAACTTCTCGGTGATCTCC 1500  
Db 1441 TAGTAGGTCCTCGATCTTGTGGCGAGTTTAGCGAGTAGCCGAACTTCTCGGTGATCTCC 1500  
QY 1501 GCAAAACCGCGATGATCTTTCGGCAGACTTTCGAAACCTTCGACAAAGTCCCGGATTTCT 1560  
Db 1501 GCAAAACCGCGATGATCTTTCGGCAGACTTTCGAAACCTTCGACAAAGTCCCGGATTTCT 1560  
QY 1561 TCTCGGTTGGTTCGACAGATCTTAACGAACTTCGGACTCTTGAATGTCATCGAA 1620  
Db 1561 TCTCGGTTGGTTCGACAGATCTTAACGAACTTCGGACTCTTGAATGTCATCGAA 1620  
QY 1621 CTTGACTCCGCTAGGCTTGTCTTATATTTTCAGGCTATCATAGTTAACTTACATACACTTA 1680  
Db 1621 CTTGACTCCGCTAGGCTTGTCTTATATTTTCAGGCTATCATAGTTAACTTACATACACTTA 1680  
QY 1681 ACTCAATAATAGGATAGATTAATTAACCCATCAATGATTTTCAATCAATAAATTCGAC 1740  
Db 1681 ACTCAATAATAGGATAGATTAATTAACCCATCAATGATTTTCAATCAATAAATTCGAC 1740  
QY 1741 ATTCAAAACATCCGTAATCAATCAACCATCAGCTATAGTTAGGTGACTATCTACTGT 1800  
Db 1741 ATTCAAAACATCCGTAATCAATCAACCATCAGCTATAGTTAGGTGACTATCTACTGT 1800  
QY 1801 GATCCGTACGTGAAGTTAGCGAGTCATGATCCAGTCTGTGACCTTATTTGGCCGAACAG 1860  
Db 1801 GATCCGTACGTGAAGTTAGCGAGTCATGATCCAGTCTGTGACCTTATTTGGCCGAACAG 1860  
QY 1861 TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCCTACCGGTCTCTTTTTTAT 1920  
Db 1861 TATCCCTTATCCAAATCCAGTCTTCTCAACTCTTCTAGCCTACCGGTCTCTTTTTTAT 1920  
QY 1921 ACTTTTGAAGAAATCAAAATCAAAACAGATCAAAATAACACGGTGAGACATGTGACAT 1980  
Db 1921 ACTTTTGAAGAAATCAAAATCAAAACAGATCAAAATAACACGGTGAGACATGTGACAT 1980  
QY 1981 GCTAGTCTCTGAAAGCAATTAATTCGCGCATCCACAGACGTCTCAGCTTCAATCAACCCAC 2040  
Db 1981 GCTAGTCTCTGAAAGCAATTAATTCGCGCATCCACAGACGTCTCAGCTTCAATCAACCCAC 2040  
QY 2041 TTTTTCCTACATACCATGTCGATGGCTTTGTTGATGACAGACCAACAGCTTGCCTT 2100  
Db 2041 TTTTTCCTACATACCATGTCGATGGCTTTGTTGATGACAGACCAACAGCTTGCCTT 2100  
QY 2101 TGGTTGTGCTTAACAG 2160  
Db 2101 TGGTTGTGCTTAACAG 2160

GENERAL INFORMATION:  
; APPLICANT: May, Gregory D.  
; APPLICANT: Clendennen, Stephanie K.  
; APPLICANT: Mason, Hugh S.  
; APPLICANT: Lim, Miguel A. Gomez  
; APPLICANT: Airtzen, Charles J.  
; TITLE OF INVENTION: DNA Regulatory Elements Associated with Fruit Development  
; FILE REFERENCE: 031998-007  
; CURRENT APPLICATION NUMBER: US/09/892,635A  
; CURRENT FILING DATE: 2001-06-28  
; PRIOR APPLICATION NUMBER: US 09/160,351  
; PRIOR FILING DATE: 1998-09-25  
; PRIOR APPLICATION NUMBER: US 60/060,062  
; PRIOR FILING DATE: 1997-09-25  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 2156  
; TYPE: DNA  
; ORGANISM: Musa acuminata  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 879  
; OTHER INFORMATION: n = A,T,C or G

US-09-892-635A-44

Query Match	99.0%	Score 2139;	DB 13;	Length 2156;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 2155;	Conservative	0; Mismatches	1; Indels	4; Gaps
1;				
QY 1	GGATCCCAACTTTT	TAGGAATGGATCTT	AAAAATTTAGTATATAGCTT	CAAAAGTTAGAAAAA 60
DB 1	GGATCCCAACTTTT	TAGGAATGGATCTT	AAAAATTTAGTATATAGCTT	CAAAAGTTAGAAAAA 60
QY 61	TCCTTACCAAGAGCTT	TGAGTCCATGATGACAT	CCCGTGAACCGGTGTACATG	CTCCGA 120
DB 61	TCCTTACCAAGAGCTT	TGAGTCCATGATGACAT	CCCGTGAACCGGTGTACATG	CTCCGA 120
QY 121	TGGACTCAGTGTGTTT	CATTTCGGAAGAGTTT	CGAAGAGTGCATGAAGATATG	ATTTTGG 180
DB 121	TGGACTCAGTGTGTTT	CATTTCGGAAGAGTTT	CGAAGAGTGCATGAAGATATG	ATTTTGG 180
QY 181	ATTCTTTCACTCGGTT	TGGTGCCTTCATGAGTGAC	CTCAAGAGTCCCTCAAAATAT	CAAAAG 240
DB 181	ATTCTTTCACTCGGTT	TGGTGCCTTCATGAGTGAC	CTCAAGAGTCCCTCAAAATAT	CAAAAG 240
QY 241	CCGAATCAGAAATG	AAATGTGATGAAATTCAT	TTTGTCTAATGCAAAAA	CAGGSCAT 300
DB 241	CCGAATCAGAAATG	AAATGTGATGAAATTCAT	TTTGTCTAATGCAAAAA	CAGGSCAT 300
QY 301	TCATAGCCTTTGTTT	TAAAGCAAAACATTT	CTCTCCGATTCATCCCATTC	CGTCAATCG 360
DB 301	TCATAGCCTTTGTTT	TAAAGCAAAACATTT	CTCTCCGATTCATCCCATTC	CGTCAATCG 360
QY 361	GAAGAGAAAATTTT	TGAAATCCATTTTCG	ACATAGACCAAAAGCT	CGAAATCCATCGATG 420
DB 361	GAAGAGAAAATTTT	TGAAATCCATTTTCG	ACATAGACCAAAAGCT	CGAAATCCATCGATG 420
QY 421	GAAATGAGGAGAT	CTCTATAGATTTT	CCAAATACATGTAAT	TCGACTCATTTAAACATA 480
DB 417	GAAATGAGGAGAT	CTCTATAGATTTT	CCAAATACATGTAAT	TCGACTCATTTAAACATA 476
QY 481	GGTGGATGTGTAAT	GAAATGACCTCATG	CSCTATCTCTCTGGG	TATTAACCAATAT 540
DB 477	GGTGGATGTGTAAT	GAAATGACCTCATG	CSCTATCTCTCTGGG	TATTAACCAATAT 536
QY 541	GAGAGTGAGCTTT	GCTCTGATACCAAT	TGTTAGGATCAGAGT	GGCGACTAAGAGAGGGGG 600
DB 537	GAGAGTGAGCTTT	GCTCTGATACCAAT	TGTTAGGATCAGAGT	GGCGACTAAGAGAGGGGG 596
QY 601	GAGTGAATTAGT	CGATGGAATTA	AAACTTAAAGTTT	AAAAATGAAATCGTAAATACGAG 660
DB 597	GAGTGAATTAGT	CGATGGAATTA	AAACTTAAAGTTT	AAAAATGAAATCGTAAATACGAG 656

QY	661	AAGATTTTCGTTT	TAATAGTAACTT	GAGTAGATGA	AAACCAAAAGTTAA	CAGTAGTGTAAA 720
DB	657	AAGATTTTCGTTT	TAATAGTAACTT	GAGTAGATGA	AAACCAAAAGTTAA	CAGTAGTGTAAA 716
QY	721	TAAACAATTCGGG	AAAGTAGA	AACTCACA	ATTCAAGGAACAT	ACCAATTTAAAGTGT 780
DB	717	TAAACAATTCGGG	AAAGTAGA	AACTCACA	ATTCAAGGAACAT	ACCAATTTAAAGTGT 776
QY	781	CGGTCAAAATGAC	CTTACATCC	ACTTGTGA	AGCCCTTTTTCGA	AGAGGTCCCAACTTCCAC 840
DB	777	CGGTCAAAATGAC	CTTACATCC	ACTTGTGA	AGCCCTTTTTCGA	AGAGGTCCCAACTTCCAC 836
QY	841	TAGCAAAATCACT	TTTGAAGGGA	AGGACAA	ATPACCTCTCTTA	CNACCTTTTACAATGGTTC 900
DB	837	TAGCAAAATCACT	TTTGAAGGGA	AGGACAA	ATPACCTCTCTCT	TACNACCTTTTACAATGGTTC 896
QY	901	ATACTCTTACAAA	TTTTTCAACG	AGAAAGAGG	GTGAACATGCA	AGCAATTTGAAACAA 960
DB	897	ATACTCTTACAAA	TTTTTCAACG	AGAAAGAGG	GTGAACATGCA	AGCAATTTGAAACAA 956
QY	961	GACTTGCTAAAG	ACTTTTGCTA	AGGCTTTTTT	TCTCAATCTATT	TGCTTCTCAAAAGTTGTA 1020
DB	957	GACTTGCTAAAG	ACTTTTGCTA	AGGCTTTTTT	TCTCAATCTATT	TGCTTCTCAAAAGTTGTA 1016
QY	1021	TTCTCTGCTGAGA	ATTTAGGGGT	ATTTATAG	ACCCCAAGAGG	ATTTAAATTTGGGTCCA 1080
DB	1017	TTCTCTGCTGAGA	ATTTAGGGGT	ATTTATAG	ACCCCAAGAGG	ATTTAAATTTGGGTCCA 1076
QY	1081	AAATTCGAATGCT	TTCCGGTTC	CCGAGGTTCC	CGGTGCCACCG	CTCTCAGTGTTTGACA 1140
DB	1077	AAATTCGAATGCT	TTCCGGTTC	CCGAGGTTCC	CGGTGCCACCG	CTCTCAGTGTTTGACA 1136
QY	1141	CTGGACAGTGTAC	TAGCGGTG	CCACCGCGG	AGACTCTTCG	GGGTGTGGCGGTGCCACCGC 1200
DB	1137	CTGGACAGTGTAC	TAGCGGTG	CCACCGCGG	AGACTCTTCG	GGGTGTGGCGGTGCCACCGC 1196
QY	1201	CTAGACTTTTCA	GCTCAGTGT	GCATTCGAA	ATTCGACCCAA	ACAGTCCGAACTCGG 1260
DB	1197	CTAGACTTTTCA	GCTCAGTGT	GCATTCGAA	ATTCGACCCAA	ACAGTCCGAACTCGG 1256
QY	1261	GTCCAATGACCG	TAACCGGAT	TAGGATTA	ACCTTAATCTT	ACCCCTTAATATATG 1320
DB	1257	GTCCAATGACCG	TAACCGGAT	TAGGATTA	ACCTTAATCTT	ACCCCTTAATATATG 1316
QY	1321	CAAACTACGCA	CACTGAAAT	ATATAGTCT	TAAAGCAAGTTT	TAAACCGGCAAAACGTCGAGTCT 1380
DB	1317	CAAACTACGCA	CACTGAAAT	ATATAGTCT	TAAAGCAAGTTT	TAAACCGGCAAAACGTCGAGTCT 1376
QY	1381	TCCTTCGGGAT	TCCTTCGGG	CAGACTTCT	GTATAPACCTT	GTGATTTCTTTAGCGGACTCC 1440
DB	1377	TCCTTCGGGAT	TCCTTCGGG	CAGACTTCT	GTATAPACCTT	GTGATTTCTTTAGCGGACTCC 1436
QY	1441	TAGTAGGTTCCG	ATCTGTGCG	AGGTTTAG	CGAGTAGCGG	ACCTTCTCGGTGATCTCC 1500
DB	1437	TAGTAGGTTCCG	ATCTGTGCG	AGGTTTAG	CGAGTAGCGG	ACCTTCTCGGTGATCTCC 1496
QY	1501	GCAAAACCGCG	ATGATCTCT	TCGGCAGACT	TTTTCGAAAA	CTTCGACAAGTCCCGGATTTCT 1560
DB	1497	GCAAAACCGCG	ATGATCTCT	TCGGCAGACT	TTTTCGAAAA	CTTCGACAAGTCCCGGATTTCT 1556
QY	1561	TCCTCGGTTGG	TTCGCGAC	AGATCTTAA	CGAAACCTTCG	GACTCTTTGAAATGTCATCGAA 1620
DB	1557	TCCTCGGTTGG	TTCGCGAC	AGATCTTAA	CGAAACCTTCG	GACTCTTTGAAATGTCATCGAA 1616
QY	1621	CTTGACTCCG	GTAGGCTCT	TATATTTT	TTCAGGCTAT	CATAGTTAATCTTACATCTTA 1680
DB	1617	CTTGACTCCG	GTAGGCTCT	TATATTTT	TTCAGGCTAT	CATAGTTAATCTTACATCTTA 1676
QY	1681	ACTCAATAAT	TATGATTA	TAATTAAC	CCCATCAAT	TGATTTTCATCATCAAAATTCGAC 1740
DB	1677	ACTCAATAAT	TATGATTA	TAATTAAC	CCCATCAAT	TGATTTTCATCATCAAAATTCGAC 1736



1261 GTCCAAATTGACCCGTAACCGGATTATAGGATTAAACCCCTTAATCCTAACCCCTAAATTATATG 1320  
1257 GTCCAAATTGACCCGTAACCGGATTATAGGATTAAACCCCTTAATCCTAACCCCTAAATTATATG 1316  
1321 CAAACTACGCAACTGAAATATAGTCTTAAGCAAGTCTTTTAAACCGGCAAAACGTCGAGTCT 1380  
1317 CAAACTACGCAACTGAAATATAGTCTTAAGCAAGTCTTTTAAACCGGCAAAACGTCGAGTCT 1376  
1381 TCTTCGGCGCATCTTTCGGCAGACTCTCGATATACCTTTGGAATTTCTTACGGGACTCC 1440  
1377 TCTTCGGCGCATCTTTCGGCAGACTCTCGATATACCTTTGGAATTTCTTACGGGACTCC 1436  
1441 TAGTAGGTCCTCGATCTTGTGGCAGATTTCGAGTAGCCGAACTCTTCCTGGTGCATCTCC 1500  
1437 TAGTAGGTCCTCGATCTTGTGGCAGATTTCGAGTAGCCGAACTCTTCCTGGTGCATCTCC 1496  
1501 GCAAAACCGCGATGATCTCTTCGGCAGACTCTTCGAAACTTCGAAAGTCCCGGATTCT 1560  
1497 GCAAAACCGCGATGATCTCTTCGGCAGACTCTTCGAAACTTCGAAAGTCCCGGATTCT 1556  
1561 TCTCGGTGGTTCGGCAGACTCTTAACGAACTTCGAAAGTCCCGGATTCT 1620  
1557 TCTCGGTGGTTCGGCAGACTCTTAACGAACTTCGAAAGTCCCGGATTCT 1616  
1621 CTTCGACTCCGCTAGGCTTGTCTTATATTTTCAGGCTATCATAGTAAATCTTACATCTTA 1680  
1617 CTTCGACTCCGCTAGGCTTGTCTTATATTTTCAGGCTATCATAGTAAATCTTACATCTTA 1676  
1681 ACTCAATTAATGATAGTAAATTAATTAACCATCAATGATTTTATCATCAATCAAAATTCGAC 1740  
1677 ACTCAATTAATGATAGTAAATTAATTAACCATCAATGATTTTATCATCAATCAAAATTCGAC 1736  
1741 ATTCAAAACATCCGTAATCAATTAACCATCAATGATTTTATCATCAATCAAAATTCGAC 1800  
1737 ATTCAAAACATCCGTAATCAATTAACCATCAATGATTTTATCATCAATCAAAATTCGAC 1796  
1801 GATCCGTAACGTAAGTTCGAGTCAATGATCCAGGTCGTCTCACTTATTTGGCCGAAACAG 1860  
1797 GATCCGTAACGTAAGTTCGAGTCAATGATCCAGGTCGTCTCACTTATTTGGCCGAAACAG 1856  
1861 TATCCCTTATCCAAATCCAGTCTTCTCAATTTAGCCTACCGGCTCTCTTTTATTTAT 1920  
1857 TATCCCTTATCCAAATCCAGTCTTCTCAATTTAGCCTACCGGCTCTCTTTTATTTAT 1916  
1921 ACTTTTGAAGAATTCGAATCAAAACAGATACAAATTAACCGTTCGATGACAT 1980  
1917 ACTTTTGAAGAATTCGAATCAAAACAGATACAAATTAACCGTTCGATGACAT 1976  
1981 GCTAGTCTCTGGAAGCATTAATTCGCGCATCCACAGACGTCGTCTCAGCTTCATCACCAC 2040  
1977 GCTAGTCTCTGGAAGCATTAATTCGCGCATCCACAGACGTCGTCTCAGCTTCATCACCAC 2036  
2041 TTTTTCCTACAT-ACCATGTCGATGCTTCTCAATCTTTCAGCCTACCGGCTCTCTTTTATTT 2099  
2037 TTTTTCCTACAT-ACCATGTCGATGCTTCTCAATCTTTCAGCCTACCGGCTCTCTTTTATTT 2096  
2100 TTGGTTGTGCTTAAACAG 2159  
2097 TTGGTTGTGCTTAAACAG 2152  
2160 G 2160  
2153 G 2153

RESULT 4  
US-10-298-122-3  
; Sequence 3, Application US/10298122  
; Publication No. US20030221214A1  
; GENERAL INFORMATION:  
; APPLICANT: Gmitter, Frederick G  
; APPLICANT: Deng, Zhanao

APPLICANT: Zhang, Hongbin  
; TITLE OF INVENTION: CITRUS TRISTEZA VIRUS AND METHODS OF USE  
; FILE REFERENCE: 5853-220  
; CURRENT APPLICATION NUMBER: US/10/298,122  
; CURRENT FILING DATE: 2002-11-15  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 78338  
; TYPE: DNA  
; ORGANISM: artificial  
; FEATURE:  
; OTHER INFORMATION: Poncirus trifoliata and citrus grandis hybrid  
US-10-298-122-3  
Query Match 5.7%; Score 122.2; DB 16; Length 78333;  
Best Local Similarity 59.8%; Pred. No. 7e-21;  
Matches 222; Conservative 0; Mismatches 148; Indels 1; Gaps 1;  
QY 6 CCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAGTTCAAAGTTAGAAAAATCTTT 65  
Db 47226 CCACTCCCTTAGGTAGAGATCTAATGATTTTCTTGACTTTTTCGCTATTGAAAAAGTCTT 47285  
QY 66 ACCAAGAGCTTTGAGTCCATTGATGACATCGGTGAAC-CGTGTACATGTCTCCGATGGA 124  
Db 47286 TCCTAGGCTCCTAGAGTGTCTACTATGTCCTGGAATCTAGTATACATAGATACACATT 47345  
QY 125 CTCACITGGTTTCATTTCGGAAGAGTTCGAAAGAGTGCATAAGAAATATGATTTTGGATTC 184  
Db 47346 TTCATTTTGTTCATTTCGGAAGAGTTCGAAAGAGTGCATAAGAAATATGATTTTGGATTC 47405  
QY 185 TTTCACTCGTTCGTCCTTCATGAGTACCTCAAGAGTCTTCAAAATATCAAAAGCCGA 244  
Db 47406 TTTCACTCGTTCGTCCTTCATGAGTACCTCAAGAGTTCGTCCTTCAATAGCACT 47465  
QY 245 ATCACAATTAAGAAATGATGATTGAATTCATTTTGTCTTAATGCAACAAACAGGCACTCAT 304  
Db 47466 TTCGCACTAGACACTCTATGAACCTCTTACTTATCTAGTGCACAGAACAGGCAITCAT 47525  
QY 305 AGCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTCGCTCATCGGAAG 364  
Db 47526 GGCCTTGAATTTAGAGAAGCTTTTCTCTATCGAATTCATTCOAATCCCGTGAAGTTT 47585  
QY 365 AGAAATTTT 375  
Db 47586 TGGATATCTT 47596

RESULT 5  
US-10-424-599-108338  
; Sequence 108338, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 108338  
; LENGTH: 2760  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_68844C.1  
US-10-424-599-108338

Query Match 3.8%; Score 81.4; DB 13; Length 2760;  
Best Local Similarity 53.8%; Pred. No. 9.7e-11;

Matches 189; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
QY 1 GGAATCCAACTTTAGGAAGTCTTAAATTTTGTATTAAGTTCAAGTTAGAAAA 60  
Db 1829 GGTGCGCATCTCTACTTAAACATCTTAAACACCTTGTATTAAAGATCTCTATTTTGAAT 1888  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTCATGACATCCCTGAAACGGTGTACATGTCCTCG- 119  
Db 1889 TCTTTGCTAAGCTGTAGATGATTTACTATATGTGTAAATCTCTTTTGCATCTCTGA 1948  
QY 120 ATGACTCACTTGTTTCATTGCGAAAAAGTTGCGAAGAGTGTCATAGAAATATTGATTTTG 179  
Db 1949 ATATTTTCATTTGATTCATCTTAAATTAATTCATCTACTCATGAGTTAGTCATTTATCCTA 2008  
QY 180 GATCTTTTCACTCGGTTGGTCCCTTCATGAGTGACCTCAAGAGTCTCCCAAAATATCAAAA 239  
Db 2009 GATCTTTTAAATTTGTAGTTCTCTTCATGTGTTAATCGAAGAGTGTCCCAATTTCTCTTA 2068  
QY 240 GCCGAATCAGAAATGAATGTGATTTGAATTCATTTTGTCTAATGACAAAAACAGGGCA 299  
Db 2069 GCATCTTACAAATTTGAACCCCTGAAATATTCATCCATCTCTAGGCGAGATGTTATTATG 2128  
QY 300 TTCATAGCTTTGTGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCAT 350  
Db 2129 TTTTGGCTTTAAGTTGATTTGACTCTTTTCTATCTCTTCAGACCAT 2179  
RESULT 6  
US-10-298-122-2  
; Sequence 2, Application US/10298122  
; Publication No. US20030221214A1  
; GENERAL INFORMATION:  
; APPLICANT: Gmitter, Frederick G  
; APPLICANT: Deng, Zhanao  
; APPLICANT: Zhang, Hongbin  
; TITLE OF INVENTION: CITRUS TRISTEZA VIRUS AND METHODS OF USE  
; FILE REFERENCE: 5853-220  
; CURRENT APPLICATION NUMBER: US/10/298, 122  
; CURRENT FILING DATE: 2002-11-15  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 58965  
; TYPE: DNA  
; ORGANISM: artificial  
; FEATURE:  
; OTHER INFORMATION: Poncirus trifoliata and Citrus grandis hybrid  
; NAME/KEY: misc feature  
; LOCATION: (2145)..(2320)  
; OTHER INFORMATION: n denotes unsequenced nucleotides  
US-10-298-122-2  
Query Match 3.8%; Score 81.2; DB 16; Length 58965;  
Best Local Similarity 57.7%; Pred. No. 8.1e-10;  
Matches 188; Conservative 0; Mismatches 128; Indels 10; Gaps 2;  
QY 181 ATTCTTCACTCGGTTGGTCCCTCATGAGTCACTCAAGAGTCTCTCAATATCAAAAG 240  
Db 38730 ATTCTCTCTCTTTGTAGTCTCTCATACACAAATTTTGAGTTTCTCCAAATTTTCATAAG 38789  
QY 241 CCGAATCAAAATTCAGATGTGATTTCAATTTTGTCTAATGCACAAAACAGGGCAT 300  
Db 38790 CATTAGACAAATTCAGAACTCTATGAAATTCCTTTTATCTAGGACACAAATAAGAT 38849  
QY 301 TCATAGCTTTTGTGTTTAAAGCAAAAACATCTCTCCGATTCATCCCATTCGCTCATCG 360  
Db 38850 TCATAGCTTTGAAATTTAAAGACATATTTTCTATCTAACTCACTTAAAGCATCTTTC 38909  
QY 361 GAAGGAAATTTTGAATCCATTTTCAGACATAGACCAAGCTCGAATCCATGCGCATG 420  
Db 38910 TAGGA-----ATGAGTAAATCATCAAAACAACCTTCCCAATTTTCATAATCTAAAGCTT 38963

QY 421 GAAATGAGGAAGATCCTCATATGAGTTTCCAAATACATGTAATTCGACTCATTAAACATA 480  
Db 38964 GTAA-----ATAAATCTCATCTTAGTTTTCATAAGGATAATCATTTCCATCTAAAAAT 39019  
QY 481 GGTGGATGTGTAATGAATGACCCCTC 506  
Db 39020 GGAAGTCTTGTGATAGATTGTCCTTC 39045  
RESULT 7  
US-09-966-881-48  
; Sequence 48, Application US/09966881  
; Patent No. US20020120960A1  
; GENERAL INFORMATION:  
; APPLICANT: Seymour, Graham  
; Bird, Colin  
; Medina-Suarez, Rosybel  
; TITLE OF INVENTION: Genetic control Of Fruit Ripening  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zeneca Ag Products Inc.  
; STREET: 1800 Concord Pike  
; CITY: Wilmington  
; STATE: DE  
; COUNTRY: USA  
; ZIP: 19850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/966,881  
; FILING DATE: 28-Sep-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/242,860  
; FILING DATE: 29-Mar-1999  
; APPLICATION NUMBER: GB 9618862.8  
; FILING DATE: 10-SEP-1996  
; APPLICATION NUMBER: GB 9708366.1  
; FILING DATE: 25-APR-1997  
; APPLICATION NUMBER: PCT/GB97/02424  
; FILING DATE: 08-SEP-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hohenschutz, Liza D.  
; REGISTRATION NUMBER: 33,712  
; REFERENCE/DOCKET NUMBER: SEE 50183/UST  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (302) 886-1699  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 809 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: U-D86  
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:  
US-09-966-881-48  
Query Match 3.4%; Score 73.4; DB 9; Length 809;  
Best Local Similarity 98.7%; Pred. No. 6.5e-09;  
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2086 CCACAAGCTTGCTTTGTTGTTGCTTAACAGAGAGAGAGAGACGACCGATAGCCT 2145  
Db 9 CCACAAGCTTGCTTTGTTGTTGCTTAACAGAGAGAGAGAGACGACCGATAGCCT 68  
QY 2146 CCTCATTCCACCATGG 2160  
Db 69 CCTCATTCCATATGG 83

RESULT 8  
 US-10-334-703-3/c  
 ; Sequence 3, Application US/10334703  
 ; Publication No. US2003015451A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Laten, Howard M.  
 ; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS OF USE THEREOF  
 ; NUMBER OF SEQUENCES: 58  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, Gerstein and Borun  
 ; STREET: 233 South Wacker Drive/6300 Sears Tower  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/334,703  
 ; FILING DATE: 20-Dec-2002  
 ; CLASSIFICATION: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wrona, Thomas J.  
 ; REGISTRATION NUMBER: 44,410  
 ; REFERENCE/DOCKET NUMBER: 27013/33214C US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (312) 474-6300  
 ; TELEFAX: (312) 474-0448  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2417 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 US-10-334-703-3

Query Match 2.9%; Score 62.8; DB 15; Length 2417;  
 Best Local Similarity 50.6%; Pred. No. 1e-05;  
 Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;  
 QY 4 TCCCACTTTTAGGAATGGATCTTAAAAATTTTAGTTTATAAGTTCAAAGTTAGAAAAATCT 63  
 Db TCAAATCTCTTAGGCAAGGATCTGAGGATCTTTCTCACCAGCTTTTCATCAGTCATTCIT 727  
 QY 64 TTACCAAGAGCTTTGAGTCAATGATGACATC-CGTGAACGGGTGACATGCTCCGATG 122  
 Db TCTCCCAAGCAGTGAAGCATTTGGCAATTTCAAGAAATGTTTCATGTGAAAGTCATGAATA 667  
 QY 123 GACTCACTTTGGTTTCATTCGAAAAGTTCGAAAGAGTGCATAAGATAATTTGATTTGGAT 182  
 Db CACTCTTCCTCTTCATCTCAGATTTTGAATTTTGCGCCAAATAGTTGCAATCTGGAC 607  
 QY 183 TCTTTCACCTGGTGTGGCTTCATGATGACCTCAAGAGTCTCCCAAAATATCAAAAGCC 242  
 Db ATCTTCACITTTGGAGGTTCCTTCATGAGTGGTTTTCAGGATCTCCCATGATCTTGGCC 547  
 QY 243 GAATCAAAATGAAATGTAATTAATTTGTTCTTAATGCACAAAACAGGGCATTC 302  
 Db ACTGTGCATGTGTGATCATGCTGAAGATATTTCTGTCAACTCCATTGAATAGAGCATTC 487  
 QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTC 352  
 Db AAAGCTTTGGAGTTTCCAAGTGCCAAATTCGTCTCTTTTAGTCCAGTC 437

RESULT 9  
 US-10-396-122-3/c  
 ; Sequence 3, Application US/10396122  
 ; Publication No. US2003022122A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Laten, Howard  
 ; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF  
 ; FILE REFERENCE: 27013/39234  
 ; CURRENT APPLICATION NUMBER: US/10/396,122  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: 2003-03-25  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 105  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 3  
 ; LENGTH: 2417  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: SIRE 1 cDNA clone  
 US-10-396-122-3

Query Match 2.9%; Score 62.8; DB 16; Length 2417;  
 Best Local Similarity 50.6%; Pred. No. 1e-05;  
 Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;  
 QY 4 TCCCACTTTTAGGAATGGATCTTAAAAATTTTAGTTTATAAGTTCAAAGTTAGAAAAATCT 63  
 Db TCAAATCTCTTAGGCAAGGATCTGAGGATCTTTCTCACCAGCTTTTCATCAGTCATTCIT 727  
 QY 64 TTACCAAGAGCTTTGAGTCCATGATGACATC-CGTGAACGGGTGACATGCTCCGATG 122  
 Db TCTCCCAAGCAGTGAAGCATTTGGCAATTTCAAGAAATGTTTCATGTGAAAGTCATGAATA 667  
 QY 123 GACTCACTTTGGTTTCATTCGAAAAGTTCGAAAGAGTGCATAAGATAATTTGATTTGGAT 182  
 Db CACTCTTCCTCTTCATCTTCAGATTTTTCGAAATTTTGTGCGCCAAATAGTTGCAATCTGGAC 607  
 QY 183 TCTTTCACCTGGTGTGGCTTCATGATGACCTCAAGAGTCTCCCAAAATATCAAAAGCC 242  
 Db ATCTTCACITTTGGAGGTTCCTTCATGAGTGGTTTTCAGGATCTCCCATGATCTTGGCC 547  
 QY 243 GAATCAAAATGAAATGTAATTAATTTGTTCTTAATGCACAAAACAGGGCATTC 302  
 Db ACTGTGCATGTGTGATCATGCTGAAGATATTTCTGTCAACTCCATTGAATAGAGCATTC 487  
 QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTC 352  
 Db AAAGCTTTGGAGTTTCCAAGTGCCAAATTCGTCTCTTTTAGTCCAGTC 437

RESULT 10  
 US-10-424-599-108650/c  
 ; Sequence 108650, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285884  
 ; SEQ ID NO 108650  
 ; LENGTH: 2801  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_69124C.1



US-10-424-599-108650	
Query Match 2.9%; Score 62.8; DB 13; Length 2801; Best Local Similarity 50.6%; Pred. No. 1.1e-05; Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;	
QY 4	TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAAATCT 63
Db	1093 TCAAAATCTTTAGCAAGGATCTGAGGATCTTTCTCCAGCTTTTTCATCAGTCATCTT 1034
QY 64	TTACCAAGAGCTTTGAGTCCATTCATGACATC-CGTGAAACGGTGTACATGTCTCCGATG 122
Db	1033 TCTCCCAGGAGTGCAGCATTTGCCAATTTCAAGATGTTCAATGTGAAAGTCATGAATA 974
QY 123	GACTCACTTGGTTTCATTCGGAAAGTTCGAAAGAGTGATCAAGAAATTTGATTTGGAT 182
Db	973 CACTCTTCCCTTCATCTTCAGATTTTCGAATTTTGGGCCAATAGTTGCAATCTGGAC 914
QY 183	TCCTTCACTCGTGTGGTCCCTCATGAGTGACCTCAAGAGTCCCTCCAAATATCAAAAGCC 242
Db	913 ATCTTCACCTTTGGAGGTTCCCTTCAGAGTGGTTTTCAGGATCTCCCATGCAATCTTGCC 854
QY 243	GAATCACAAAATGAAATGTGAATTCATTTTGTCTAATGACAAAAACAGGGCATTC 302
Db	853 ACTGTGATGTGTTGATCAGTCTGAAGATATTTCTGTCACTTCCATTCGAATAGAGCATTC 794
QY 303	ATAGCCTTTGGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCATTC 352
Db	793 AAAGCTTTGGAGTTTCCAAAGTCCAAATTCGTCTTCTTCTTTAGTCCAGTC 744
RESULT 11	
US-10-396-122-93/c	
; Sequence 93, Application US/10396122	
; Publication No. US2003021222A1	
; GENERAL INFORMATION:	
; APPLICANT: Laten, Howard	
; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF	
; FILE REFERENCE: 27013/39234	
; CURRENT APPLICATION NUMBER: US/10/396,122	
; CURRENT FILING DATE: 2003-03-25	
; PRIOR APPLICATION NUMBER: US 60/367,302	
; PRIOR FILING DATE: 2002-03-25	
; NUMBER OF SEQ ID NOS: 105	
; SOFTWARE: PatentIn version 3.0	
; SEQ ID NO 93	
; LENGTH: 9399	
; TYPE: DNA	
; ORGANISM: Glycine max	
; FEATURE:	
; NAME/KEY: misc_feature	
; OTHER INFORMATION: Soybean retroelement SIRE1 9	
; FEATURE:	
; NAME/KEY: CDS	
; LOCATION: (1215)..(5942)	
; OTHER INFORMATION: gag protease reverse transcriptase RNAase H polypeptide	
; FEATURE:	
; NAME/KEY: CDS	
; LOCATION: (5946)..(7985)	
; OTHER INFORMATION: envelope protein	
; FEATURE:	
; NAME/KEY: LTR	
; LOCATION: (29)..(1156)	
; FEATURE:	
; NAME/KEY: LTR	
; LOCATION: (8606)..(9361)	
US-10-396-122-93	
Query Match 2.9%; Score 62.8; DB 16; Length 9399; Best Local Similarity 50.6%; Pred. No. 2.5e-05; Matches 177; Conservative 0; Mismatches 172; Indels 1; Gaps 1;	
QY 4	TCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATAAGTTCAAAGTTAGAAAAATCT 63



Db 5104 GTCTCAAGGTTGCTCACTCTATCAAACTCCCTCTGGCTCAAGAGATTGAACAAGGCATTC 5163  
QY 303 AT 304  
Db 5164 AT 5165

## RESULT 13

US-10-396-122-87/c  
; Sequence 87, Application US/10396122  
; Publication No. US20030221222A1  
; GENERAL INFORMATION:  
; APPLICANT: Laten, Howard  
; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF  
; FILE REFERENCE: 27013/39234  
; CURRENT APPLICATION NUMBER: US/10/396,122  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: US 60/367,302  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 105  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 87  
; LENGTH: 9072  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: SIRE1 7  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1264)..(5991)  
; OTHER INFORMATION: gag protease integrase reverse transcriptase RNAase H polyprotein  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (5995)..(8040)  
; OTHER INFORMATION: envelope protein  
; FEATURE:  
; NAME/KEY: LTR  
; LOCATION: (58)..(1205)  
; FEATURE:  
; NAME/KEY: LTR  
; LOCATION: (8677)..(9072)  
US-10-396-122-87

Query Match 2.8%; Score 59.6; DB 16; Length 9072;  
Best Local Similarity 50.0%; Pred. No. 0.00018;  
Matches 175; Conservative 0; Mismatches 174; Indels 1; Gaps 1;

QY 4 TCCCACTTTTAGGAATGGATCTTAAATTTTAGTTTATAAGTTCAAAGTTAGAAAAATCT 63  
Db 1814 TCAAAATCTCTTAGGCAAGGATCTGAGGATCTTTCTCACCAGCTTTTCATCTGTTATCCCTC 1755  
QY 64 TTACCAAGGCTTTGAGTCCATTTGATGACATC-CGTGAACGGGTGACATGCTCCGATG 122  
Db 1754 TCTCCCAAGCGAGTGAAGCAATTTGCAATTTTCAAGAAATGTTCAATGCGAAGTCAATGAATA 1695  
QY 123 GACTCACTTGGTTTCAATTCGAAAAAGTTTCAAGAGTGCATAGAATATTTGATTTGGAT 182  
Db 1694 CACTCTTCCTCTCATCTTCAGATTTTCAATTTTGTAGCCAAAGTTGCAATCTGGAC 1635  
QY 183 TCTTTCACTCGGTTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCC 242  
Db 1634 ATCTTCACATTTGGAGTTCTTCATGAGTCTTTTCAGGATCTCGATGCACTTTGGCC 1575  
QY 243 GAATCAAAATGAATGATGAATTCATTTTGTCTAAATGCACAAAACAGGCGATTC 302  
Db 1574 ACTGTCCAAGTGTGTGATGATCTGAAGATGTTCTTGTCAACTCCATTGAATAGTGCATTC 1515  
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTC 352  
Db 1514 AAGCTTTGGAGTTTCCAAAGTGCCAAATCGTCTCTTCTTTTGTAGTCCAGTC 1465

## RESULT 14

US-10-437-963-10782/c  
; Sequence 10782, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 10782  
; LENGTH: 5016  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_1706C.1  
US-10-437-963-10782

Query Match 2.8%; Score 59.4; DB 17; Length 5016;  
Best Local Similarity 48.9%; Pred. No. 0.00014;  
Matches 251; Conservative 1; Mismatches 247; Indels 14; Gaps 3;

QY 4 TCCCACTTTTAGGAATGGATCTTAAATTTTAGTTTATAAGTTCAAAGTTAGAAAAATCT 63  
Db 500 TCATACTTCTCCGGTAGGGCTCTAAGCATCTTTTGGGCAACCTCAAGATCGGTCTAGTTT 441  
QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCGGTGAAC-CGGTGTACATGTTCCCGATG 122  
Db 440 GCCCAAGCCCTTGAGATCATTCAAATGACATTTGAGCCTCCCATACATGTCAATCACA 381  
QY 123 GACTCACTTGGTTTCAATTCGAAAAAGTTTCAAGAGTGCATAAGAAATATTGATTTGGAT 182  
Db 380 CTCTCATGAGGCAACATGGAGAATGCTCATATTGGATCTTGAGGAATGAAGCTTTGGCA 321  
QY 183 TCTTCACTCGGTTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAAGCC 242  
Db 320 TCTTGTACTCACTTTGATCCCTCATGGATCTCCGCCAACTTGTTCCAAATCTCATATCG 261  
QY 243 GAATCAAAATTCGAAATGATTTGAATTCATTTTCTTAATGCACAAAACAGGCGATTC 302  
Db 260 GTCTCAAGGTTGCTCACTCTATCGAACTCTCTTTGGCTTCAAGAGTTGAACAAGGCATTC 201  
QY 303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTTCCGATTCATCCCATTCGCTCATCGGA 362  
Db 200 ATGGC-----TTGGCAATTGAGTTGAGGTTACGGTGATCACTCTCCGTCAGGGT 150  
QY 363 AGAGAAAAATTTTGAATCCATTTTCGAATAGACCAAGCTCGAAATTCATGCAATGA 422  
Db 149 GTGCCGGTGATATCAAGGCTTACATCCCAATATCTCCAAATATGGAA-----GCTCATAGC 94  
QY 423 AATGAGGAAGATCCTCATATGATTTTCCAAATACATGTAATTCGACTCATTAACACATAG 482  
Db 93 TTTGAGGTGAGTAGACATTTTAAATTTTCCAGAGTAGTTTGTGCTCCATTAAACATGA 34  
QY 483 TGGATGTGTAATGAATGACCCCTCATGCSCTAT 515  
Db 33 AGCCTCCCTACATGGTTCCACCTCGTTCGACAT 1

## RESULT 15

US-10-437-963-44132/c  
; Sequence 44132, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:

Mon Sep 20 12:58:19 2004

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 44132
; LENGTH: 5304
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_47222C.1
US-10-437-963-44132

Query Match          2.8%; Score 59.4; DB 17; Length 5304;
Best Local Similarity 48.9%; Pred. No. 0.00014;
Matches 251; Conservative 1; Mismatches 247; Indels 14; Gaps 3;

Qy      4 TCCCAACTTTTAGAATGGATCTTAAATTTTATAGTTTCAAGTTAGAAAAATCT 63
Db      |||
Qy      500 TCATACCTTCGGGTAGGGCTCTAAGCATCTTTGGGCAACCTCAAGATCGGTAGTTT 441
Db      |||
Qy      64 TTACCAAGAGCTTTGAGTCCATGATGACATCCGCGAAC--GGGTACATGCTCCGATG 122
Db      |||
Qy      440 GCCCAAGCCCTTGAGATCATTCACATGATGAGCCCTCCCATCATGTCATTCACA 381
Db      |||
Qy      123 GACTCACCTTGTTTCATTCGGAAAAGTTCGAAAGAGTGATAGATAATTCATTTTGGAT 182
Db      |||
Qy      380 CTCATGAGGACATCGAGATGCTCTCATATTGGATCTTGAGGAATGAAGCTTGGA 321
Db      |||
Qy      183 TCTTTCACGCGTGGTCCCTTCATGAGTGACCTCAAGAGTCCCTCCAAATATCAAAGCC 242
Db      |||
Qy      320 TCCTTGTACTCACTTGTAACCTCATGGATCTCCGCCAATCTGTTCCAAATCTCATGCG 261
Db      |||
Qy      243 GAATCACAAATGAATGTGATTTGATTCATTTTGTCTAATGCACAAAACAGGCGATTC 302
Db      |||
Qy      260 GTCTCAAGGTGCTCCTCATCGAACTCCTCTTGGCTCAAGAGTTGAACAAGGCATTC 201
Db      |||
Qy      303 ATAGCCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCATTCGCTCATCGGA 362
Db      |||
Qy      200 ATGGC-----TTGGGCATTGAGTTGGAGTTACGGTGATCACTCTCCGTCAAGGGT 150
Db      |||
Qy      363 AGAGAAATTTTGAATTCATTTTCGACATAGACCAAGCTCGAAATCCATGCATGGA 422
Db      |||
Qy      149 GTGCCGGTGATATCAAGCCCTACATCCCAATCTCCAAATATGGAA---GCTCATAGC 94
Db      |||
Qy      423 AATGAGGAAGATCCCTCATATGAGTTTCCATATACATGTAATTCGACTCATTAACATAGG 482
Db      |||
Qy      93 TTTGAGGTGAGTAGACATTTTAATTTCCAGTGGAGTAGTTTGTGCCATTAAACATGGA 34
Db      |||
Qy      483 TGGATGTGTAATGAATGACCTCATGCSCTAT 515
Db      |||
Qy      33 AGCCTTCCTACATGGTTCAACCTCGTTCGACAT 1
Db      |||
```

Search completed: September 18, 2004, 18:55:10  
Job time : 971.897 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 17:55:10 ; Search time 4973.1 Seconds  
(without alignments)  
12970.230 Million cell updates/sec

Title: US-09-892-635A-45  
Perfect score: 2160  
Sequence: 1 ggatcccaacttttagaat.....agctctcattcaccatgg 2160

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_estc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_hum:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gsl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	110.4	5.1	500	14	CD474210 nad03-10m
2	96.6	4.5	790	28	BH021865 GH_MBB000
c 3	92.6	4.3	977	29	CG823740 SOYDB38TH
	92	4.3	832	29	CG818965
					CG818965 SOYED06TV

C 5	91.4	4.2	821	29	CG817604
C 6	89.8	4.2	776	29	CG822306
C 7	83	3.8	732	28	BH794864
C 8	83	3.8	791	29	CG824360
C 9	81.8	3.8	890	29	CG817563
C 10	81.8	3.8	927	29	CG820967
C 11	81.4	3.8	733	29	CG815315
C 12	81.4	3.8	872	29	CG817875
C 13	81.2	3.8	295	14	CB340076
C 14	81.2	3.8	795	29	CG816026
C 15	81.2	3.8	914	29	CG822083
C 16	80.6	3.7	964	28	CC412052
C 17	79.8	3.7	396	29	AG336150
C 18	79.8	3.7	746	29	CG821955
C 19	79.8	3.7	760	29	CG817507
C 20	79.8	3.7	862	29	CG813886
C 21	79.4	3.7	699	12	BI698341
C 22	79	3.7	750	29	CG816271
C 23	78.2	3.6	607	14	CF920770
C 24	78.2	3.6	709	29	CG820051
C 25	78.2	3.6	846	29	CG819170
C 26	78.2	3.6	941	29	CG818537
C 27	78.2	3.6	986	29	CG886466
C 28	77.4	3.6	787	29	CC804368
C 29	77.4	3.6	845	28	BZ648290
C 30	76.8	3.6	352	29	CG816199
C 31	76.8	3.6	722	12	BI678830
C 32	76.8	3.6	803	28	BZ966238
C 33	76.6	3.5	608	28	BZ778766
C 34	76.6	3.5	860	29	CG823490
C 35	76.4	3.5	767	29	CG795738
C 36	76.4	3.5	792	29	CG900423
C 37	76.4	3.5	970	29	CG878233
C 38	76.2	3.5	873	29	CG227901
C 39	75.8	3.5	707	28	BZ527252
C 40	75.8	3.5	856	29	CG734878
C 41	75.4	3.5	383	28	BH780259
C 42	75	3.5	283	14	CB340514
C 43	75	3.5	770	29	CG824885
C 44	75	3.5	787	29	CG678004
C 45	75	3.5	826	29	CG677989

## ALIGNMENTS

RESULT 1	CD474210	500 bp	mrna	linear	EST 04-JUN-2003
LOCUS	CD474210	nad03-10ms1-g09	Nuphar advena	cdna	clone nad03-10ms1-g09 5',
DEFINITION	CD474210	mrna sequence.			
ACCESSION	CD474210				
VERSION	CD474210.1	GI:31395478			
KEYWORDS	EST.				
SOURCE	Nuphar advena				
ORGANISM	Nuphar advena				
REFERENCE	1	(bases 1 to 500)			
AUTHORS	dePamphilis, C., Carlson, J., Ma, H., Frohlich, M., Tanksley, S., Leebens-Mack, J., Field, D., Arrington, J., Zahn, L., Kong, H., Druckemiller, M., Landherr, L., Hu, Y., Ilut, D., Wall, K., Plock, S., Chioresan, S., Albert, V., Doyle, J., Miller, W., Oppenheimer, D., Soltis, D., Soltis, P. and Theissen, G.				
	Generation of ESTs from early flower buds of Nuphar advena				
	Unpublished (2002)				
	Contact: Claude dePamphilis or James Leebens-Mack				
	Mueller Laboratory				
	Penn State University				
	208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn				
	State University, University Park, PA 16802, USA				
	Tel: 814 863 6413				



[illegible]



Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
The Center of Excellence in Soybean Research, Teaching and  
Outreach, Southern Illinois University at Carbondale and Plant  
Genomics, The Institute for Genomic Research  
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,  
USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 618 453 1797  
Email: g4082@siu.edu, jleshultz@siu.edu, cdtown@tigr.org (URL:  
http://bioinformatics.siu.edu)  
Clones approximating a minimum tiling path were re-arrayed from the  
library master plates prior to sequencing.  
For purposes of clone identification each clone name is a  
concatenation of the original clone location and its new location  
in the re-arrayed sequencing plates.  
Seq primer: GTAATACGACTCACTATAGGCG  
Class: BAC ends

# FEATURES

source  
1. .776  
/organism="Glycine max"  
/mol\_type="genomic DNA"  
/cultivar="Forrest"  
/db\_xref="taxon:3847"  
/clone="H45P23:WTP15C17"  
/clone\_lib="LargeInsertSoybeanGenLib"  
/notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 EcorI  
clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
cv. Forrest seeds were grown in greenhouse for fourteen  
days. Nuclei were isolated and embedded in agarose,  
restriction digested with Hind III BamHI or EcoRI, large  
size DNA fragments were ligated in vector V41 (pCLD04541)  
and electro transformed in DH10a cells. About 90,000  
clones from BAC libraries were fingerprinted with HindIII  
and Hae III. Version 2 (automatic build) Contigs were  
built from 78,001 fingerprints. Contigs were manually  
examined to find the best non redundant tile path through  
the contigs representing about 13,000 clones. The clones  
were end sequenced."

# ORIGIN

Query Match 4.2%; Score 89.8; DB 29; Length 776;  
Best Local Similarity 55.2%; Pred. No. 9.9e-12;  
Matches 195; Conservative 0; Mismatches 157; Indels 1; Gaps 1;  
QY 1 GGATCCCACTTTTAGGAATGGATCTTAAATTTTATAGTTCAAAAGTTAGAAAA 60  
Db |||||  
568 GGTTCGCAATCTCTACTTAACATCTTAACATTTATTTATGAGATCTGTTTGAAA 509  
QY 61 TCTTTACCAAGACCTTTGAGTCCATTCATGACATCCGCGAAAC-GGTGTACATGTCCTCG 119  
Db |||||  
508 GTTCTTCTTAATGATCAAGATGATTAACTATATGTGTGAATCTTTCTGATATCTGT 449  
QY 120 ATGGACTCACTTGGTTTCATTCGGAAAAGTTCGAAAGATGCAATAAGATTTGATTG 179  
Db |||||  
448 ATACTTTCAATTTGCTTCATCTTAATAATATCATCTCATGATTAAGTATTATCTTA 389  
QY 180 GATCTTTCACTCGTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAA 239  
Db |||||  
388 GATCTTTGACATCACTGTTTCCTCATGTGTAATCTTTAGATGTCCTCCATATCTTA 329  
QY 240 GCGGAATCAAAATTGAATGTGAATTCATTTTGTCTAATGCCAACAACAGGGCA 299  
Db |||||  
328 GCACCTTTCAATTTGAAACCCCTTAAATAATTCATCCATTCCTAGGCGAGAAGTAATA 269  
QY 300 TCCATAGCCTTTGTTTAAAGCAAAAACATCTTCTCCGATTCATCCCATTC 352  
Db |||||  
268 TTTTAGCCCTTTAAATATATACCTGCACTAATCTTTCTTCATCTTCAGACCATC 216

# RESULT 7

BH794864/c 732 bp DNA linear GSS 02-APR-2002  
LOCUS  
DEFINITION  
ME\_MBa0003J13f, Manihot esculenta Manihot esculenta genomic clone  
ME\_MBa0003J13f, genomic survey sequence.

# ACCESSION

BH794864  
BH794864.1 GI:19893414  
GSS.  
Manihot esculenta (cassava)  
Manihot esculenta  
Manihot esculenta  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids 1; Malpighiales; Euphorbiaceae; Crotonoideae;  
Manihoteae; Manihot.  
1 (bases 1 to 732)  
Tomkins,J.P., Fregene,M., Main,D., Goicoechea,J.L., Blackmon,B.,  
Atkins,M., Tohne,J. and Wing,R.A.  
New Genomic Resources for Cassava (Manihot esculenta): Development  
of a Deep-Coverage BAC Library and Preliminary STC Analysis  
Unpublished (2002)  
Contact: Tomkins J  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson University, Clemson, SC 29634, USA  
Tel: 864 656 6419  
Fax: 864 656 4293  
Email: jtmkns@clemson.edu  
Total High Quality bases = 283  
Seq primer: TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence start: 67  
High quality sequence stop: 714.  
Location/Qualifiers  
1. .732  
/organism="Manihot esculenta"  
/mol\_type="genomic DNA"  
/strain="MECW72"  
/db\_xref="taxon:3983"  
/clone="ME\_MBa0003J13f"  
/tissue\_type="Leaf"  
/lab\_host="E. coli"  
/clone\_lib="Manihot esculenta"  
/note="Vector: pCUGIBAC-1; Site 1: HindIII; Site 2: NotI;  
For more details on library preparation and sequence  
analysis see  
http://www.genome.clemson.edu/projects/stc/cassava/ME\_MBa  
To order clones from this library see  
http://www.genome.clemson.edu/orders "

# ORIGIN

Query Match 3.8%; Score 83; DB 28; Length 732;  
Best Local Similarity 57.1%; Pred. No. 5.3e-10;  
Matches 189; Conservative 0; Mismatches 140; Indels 2; Gaps 2;  
QY 4 TCCCACTTTTAGGAATGGATCTTAAATTTTATAGTTTAAAGTTAGAAAAATCT 63  
Db |||||  
452 TCTCAAGATTTTGACAGTGTCTTAAATTTTCTTCACTAGCGTAGCTTCTTCAAAAGTTT 393  
QY 64 TTACCAAGAGCTTTGAGTCCATTTGATGACATCCGTGAA-ACGGTGTACATGTCCTCGATG 122  
Db |||||  
392 TTCCTAGTAGCTTTTAGAGATTAGTAGATTAAATATATAGTACTATTTCACACCA 333  
QY 123 GACTCATTGGTTTCAT-TCGGAAGAGTTTCGAAAGAGTGCATAAGAATATTGATTTTGA 181  
Db |||||  
332 GTTTCGCCTGATTTTCATCTCAATAAATTAGAATTCAIGACCCAGCAAGTTGCTTTGA 273  
QY 182 TTCCTTTCATCGTTGGTCCCTTCATGATGACCTCAAGAGTCCCTCCAAATATCAAAAGC 241  
Db |||||  
272 TTCCTTTCCTGGATTGGGGTATTTCATAAGTCACTTGGAGTTTATCCCATTTCTTTGGT 213  
QY 242 CGAATCACAAATTGAAATGTGATTGCAATTTTGTCTATGTCACACAAACAGGCATT 301  
Db |||||  
212 GGACTCCCACTTAAACAGGATTATATTCAGTTCATCCAAAGCCCAATTAAGTATTT 153  
QY 302 CATAGCCTTTGTTTAAAGCAAAAACATTC 332  
Db |||||  
152 TAAACCTTTAGCATTGGTGGAAATATTTTC 122

```

RESULT 8
CG824360/c
LOCUS
DEFINITION
    791 bp DNA linear GSS 18-NOV-2003
    SOYD062TH LargeInsertSoybeanGenLib Glycine max genomic clone
    H31P06:WTFP13K3, genomic survey sequence.
ACCESSION
CG824360
VERSION
CG824360.1 GI:38298587
KEYWORDS
GSS.
SOURCE
Glycine max (soybean)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 791)
Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,
Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.
End sequencing of BACs comprising a provisional minimal tiling path
from a fingerprint physical map of soybean (Glycine max) cultivar
Forrest
JOURNAL
Unpublished (2003)
COMMENT
Other GSSs: SOYD062TV
Contact: Chris Town, J. L. Shultz and D. A. Lightfoot
The Center of Excellence in Soybean Research, Teaching and
Outreach, Southern Illinois University at Carbondale and Plant
Genomics, The Institute for Genomic Research
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,
USA and 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 618 453 1797
Email: ga4082@siu.edu, jlshultz@siu.edu, cdtown@igr.org (URL:
http://bioinformatics.siu.edu)
Clones approximating a minimum tiling path were re-arrayed from the
library master plates prior to sequencing.
For purposes of clone identification each clone name is a
concatenation of the original clone location and its new location
in the re-arrayed sequencing plates.
Seq primer: AATTACCTCATTAAAGG
Class: BAC ends.

FEATURES
    source
        Location/Qualifiers
            1..791
                /organism="Glycine max"
                /mol_type="genomic DNA"
                /cultivar="Forrest"
                /db_xref="taxon:3847"
                /clone="H31P06:WTFP13K3"
                /clone_lib="LargeInsertSoybeanGenLib"
                /notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 EcorI
                clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)
                cv. Forrest seeds were grown in greenhouse for fourteen
                days. Nuclei were isolated and embedded in agarose,
                restriction digested with Hind III BamHI or EcoRI, large
                size DNA fragments were ligated in vector V41 (pCLD04541)
                and electro transformed in DH10a cells. About 90,000
                clones from BAC libraries were fingerprinted with HindIII
                and Hae III. Version 2 (automatic build) Contigs were
                built from 78,001 fingerprints. Contigs were manually
                examined to find the best non redundant tile path through
                the contigs representing about 13,000 clones. The clones
                were end sequenced."

ORIGIN
Query Match 3.8%; Score 83; DB 29; Length 791;
Best Local Similarity 54.1%; Pred. No. 5.3e-10;
Matches 190; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

QY 1 GGATCCCAACTTTTAGGAATGCATCTTAAATTTTAGTTATATAGTTCAAGTTAGAAAA 60
Db 603 GTTCCCATCTCTACTTAAACATCTTACACCTTGTTTATAGATCTTCATTTGAAT 544

QY 61 TCTTTACCAAGACCTTGGATGCCATTTGATGACATCCGTTGAACCGGTGTACATGTCGCG- 119
Db 543 TCTTTGCCTAAGGCTGCTAGATGATTACTATATGTGTAATCTCTTTTGCATACTCTGA 484

```

```

QY 120 ATGGACTCACTTGTTTCATTGCGAAAAAGTTTCGAAAAGAGTGTCATAAGAAATATTGATTG 179
Db 483 ATATTTTCATTTGCAITTCATTTAAATAATTCATCTACTCATGAGTTAGTGCATTTATCTTA 424
QY 180 GATTCCTTTCACTCGGTTGGTTCCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAA 239
Db 423 GATCTTTTAACATCTGTAGTTCTCTTCATGTGTTTAATCGAAGAGTGTCCACATTTCTCTTA 364
QY 240 GCCGAATCAAAATTTGAATGTGATTGAATTCATTTTGTCTTAATGCACAAACAGGGCA 299
Db 363 GCATCTCTTACATTTTGAACCCCTGAAGTATTCATCTCCTAGTGGCAGATGTTATTATG 304
QY 300 TTCATAGCCTTTGTGTTTAAAGCAAAACATTTCTCCGATTCATCCCCAT 350
Db 303 TTTTGGCTTTTAGTTGTTATTGTTACTCGTTTCTATCCTCTTCAGACCAT 253

RESULT 9
CG817563
LOCUS
DEFINITION
    890 bp DNA linear GSS 18-NOV-2003
    SOYBC57TV LargeInsertSoybeanGenLib Glycine max genomic clone
    B50J09:MTF7J17, genomic survey sequence.
ACCESSION
CG817563
VERSION
CG817563.1 GI:38272458
KEYWORDS
GSS.
SOURCE
Glycine max (soybean)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1 (bases 1 to 890)
Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,
Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.
End sequencing of BACs comprising a provisional minimal tiling path
from a fingerprint physical map of soybean (Glycine max) cultivar
Forrest
JOURNAL
Unpublished (2003)
COMMENT
Other GSSs: SOYBC57TH
Contact: Chris Town, J. L. Shultz and D. A. Lightfoot
The Center of Excellence in Soybean Research, Teaching and
Outreach, Southern Illinois University at Carbondale and Plant
Genomics, The Institute for Genomic Research
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,
USA and 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 618 453 1797
Email: ga4082@siu.edu, jlshultz@siu.edu, cdtown@igr.org (URL:
http://bioinformatics.siu.edu)
Clones approximating a minimum tiling path were re-arrayed from the
library master plates prior to sequencing.
For purposes of clone identification each clone name is a
concatenation of the original clone location and its new location
in the re-arrayed sequencing plates.
Seq primer: GTAATACGACTCACTATAGGCG
Class: BAC ends.

FEATURES
    source
        Location/Qualifiers
            1..890
                /organism="Glycine max"
                /mol_type="genomic DNA"
                /cultivar="Forrest"
                /db_xref="taxon:3847"
                /clone="B50J09:MTF7J17"
                /clone_lib="LargeInsertSoybeanGenLib"
                /notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 EcorI
                clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)
                cv. Forrest seeds were grown in greenhouse for fourteen
                days. Nuclei were isolated and embedded in agarose,
                restriction digested with Hind III BamHI or EcoRI, large
                size DNA fragments were ligated in vector V41 (pCLD04541)
                and electro transformed in DH10a cells. About 90,000
                clones from BAC libraries were fingerprinted with HindIII
                and Hae III. Version 2 (automatic build) Contigs were

```



built from 78,001 fingerprints. Contigs were manually examined to find the best non redundant tile path through the contigs representing about 13,000 clones. The clones were end sequenced."

## ORIGIN

```

Query Match      3.8%; Score 81.8; DB 29; Length 890;
Best Local Similarity 53.8%; Pred. No. 1.1e-09;
Matches 190; Conservative 0; Mismatches 162; Indels 1; Gaps 1;

QY 1 GGATCCCACTTTAGGAATGGATCTTAAATTTAGTTATAGTTCAAGATTAGAAAAA 60
Db 412 GGTTCACCAAGTTCTTGAAGATAGTTGAGGATTTGATGTTAGCTCATCGGTTCAAA 471
QY 61 TCTTTACCAAGAGCTTTGATGTCATTCATGATCAGATCCGTGAAC-GGTGTACATGTCCTCG 119
Db 472 ATTTTCCAGACCAAGAAGTGAATTCACATATGGGTGAATCTGTTGAAGTTCTAAA 531
QY 120 ATGGAATCACTGGTTTCATTCGGAAGAGTTCGAAGAGTGCATAAGAAATATTGATTTG 179
Db 532 ATGTTTCTCCATTTCTTCAATTCGAAAGCTTCATATTGGATAGATAGTGTGTTCTT 591
QY 180 GATCTTTCATCGGTGGTGGCTTCATGAGTGACCTCAAGAGTCTCCAAATATCAAAA 239
Db 592 GATCTTCTTACATCTCTAGAGCCCTTCATGAGTGACTTCAAGCATTTTCCATATTTCTTT 651
QY 240 GCCGAATCACAATTTGAATGTGATTTGATTTGTTCTAATGCACAAAACAGGGCA 299
Db 652 GCATTTTGACCTTCAGTATGAAGATTCATCAGAGATTAACCAAGATTAAATG 711
QY 300 TCCATAGCCTTTGTTTAAAGCAAAAACATTTCTCCGATTCATCCCATTC 352
Db 712 TTTTAGCCTTTTGATCATCTTTAACTATTCTTTCTCATCATCTTTCATC 764

```

RESULT 10  
CG820967/c

LOCUS CG820967 927 bp DNA linear GSS 18-NOV-2003  
DEFINITION SOYCD687V LargeInsertSoybeanGenLib Glycine max genomic clone  
H05C12-MTP8L16, genomic survey sequence.

ACCESSION CG820967

VERSION CG820967.1

KEYWORDS GI:38279201

SOURCE GSS.

ORGANISM Glycine max (soybean)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 927)

Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J., Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.  
End sequencing of BACs comprising a provisional minimal tiling path from a fingerprint physical map of soybean (Glycine max) cultivar Forrest

Unpublished (2003)

Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
The Center of Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University at Carbondale and Plant Genomics, The Institute for Genomic Research  
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415, USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 618 453 1797  
Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL: http://bioinformatics.siu.edu)

Clones approximating a minimum tiling path were re-arrayed from the library master plates prior to sequencing.  
For purposes of clone identification each clone name is a concatenation of the original clone location and its new location in the re-arrayed sequencing plates.

Seq primer: GTAATACGACTCACTATAGGGC

Class: BAC ends.

Location/Qualifiers

## FEATURES

## source

```

1. 927
/organism="Glycine max"
/mol_type="genomic DNA"
/cultivar="Forrest"
/db_xref="taxon:3847"
/clone="H05C12-MTP8L16"
/clone_lib="LargeInsertSoybeanGenLib"
/notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 EcorI clones); Site 1: BatXi; Soybean (Glycine max (L.) Merr.) cv. Forrest seeds were grown in greenhouse for fourteen days. Nuclei were isolated and embedded in agarose, restriction digested with Hind III BamHI or EcoRI, large size DNA fragments were ligated in vector v41 (pCLD04541) and electro transformed in DH10a cells. About 90,000 clones from BAC libraries were fingerprinted with HindIII and Hae III. Version 2 (automatic build) Contigs were built from 78,001 fingerprints. Contigs were manually examined to find the best non redundant tile path through the contigs representing about 13,000 clones. The clones were end sequenced."

```

## ORIGIN

```

Query Match      3.8%; Score 81.8; DB 29; Length 927;
Best Local Similarity 55.0%; Pred. No. 1.1e-09;
Matches 181; Conservative 0; Mismatches 147; Indels 1; Gaps 1;

QY 23 ATCTTAAATTTTATAGTTTCAAGTTTCAAGATAAATCTTTTACCAAGAGCTTTGAGTC 82
Db 579 ATCTTAAATCTTTGTTTATGAGATCTTCATTTTGAAATCTTTCCCTAAGGCTGCAAGAT 520
QY 83 CATTGATGATCCGCGTGAAC-GGTGTACATCTCTCCGATGAGTCACTGTTTCATTC 141
Db 519 GATTTACTATATGTGTAAACCTCTTTTGCATGCTTTGAAATGCTTTTCATTTGCAATTC 460
QY 142 GGAAGAAGTTTGAAGAGTGCATAGAATATTGATTTTGGATCTTTTCATCGTGGTGC 201
Db 459 TAAATAGTTTCATATTCATGAGTTAATGTGTTTATCCTAGATCTTTTAAATCTATTGTC 400
QY 202 CTTTATGAGTGACCTCAAGAGTCCCTCAAAATATCAAAAGCCGAATCACAATTTGAAATGT 261
Db 399 CTTTATGTTTACTTGTAGAGTGTCACACTTCTTTCAGTCTTTCATTTGAAACCC 340
QY 262 GATTGAATTCATTTTGTCTTATGACAAAACAGGGCATTCATAGCTTTGTTTAAAG 321
Db 339 TGAATATTCATCCATTCCTAGGAGATGTAATATGTTTGGCTTTTAAATATATT 280
QY 322 CAAAACATTTCTTCCGATTCATCCCAT 350
Db 279 TTAATCGTCTTCTATACTCTTCAGACCAT 251

```

## RESULT 11

CG815315/c

LOCUS

DEFINITION

CG815315

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max (soybean)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 733)

Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J., Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.

End sequencing of BACs comprising a provisional minimal tiling path from a fingerprint physical map of soybean (Glycine max) cultivar Forrest

Unpublished (2003)

Other\_GSSs: SOYED29TH

Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
The Center of Excellence in Soybean Research, Teaching and  
Outreach, Southern Illinois University at Carbondale and Plant  
Genomics, The Institute for Genomic Research  
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,  
USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 618 453 1797  
Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL:  
http://bioinformatics.siu.edu)  
Clones approximating a minimum tiling path were re-arrayed from the  
library master plates prior to sequencing.  
For purposes of clone identification each clone name is a  
concatenation of the original clone location and its new location  
in the re-arrayed sequencing plates.  
Seq primer: GTAATACGACTCACTATAGGCG  
Class: BAC ends.

FEATURES  
source  
1..733  
/organism="Glycine max"  
/mol\_type="genomic DNA"  
/cultivar="Forrest"  
/db\_xref="taxon:3847"  
/clone="H50C23:WTP15F10"  
/clone\_lib="LargeInsertSoybeanGenLib"  
/notes="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 BcoRI  
clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
cv. Forrest seeds were grown in greenhouse for fourteen  
days. Nuclei were isolated and embedded in agarose,  
restriction digested with Hind III BamHI or EcoRI, large  
size DNA fragments were ligated in vector v41 (pCLD04541)  
and electro transformed in DH10a cells. About 90,000  
clones from BAC libraries were fingerprinted with HindIII  
and Hae III. Version 2 (automatic build) Contigs were  
built from 78,001 fingerprints. Contigs were manually  
examined to find the best non redundant tile path through  
the contigs representing about 13,000 clones. The clones  
were end sequenced."

ORIGIN

Query Match 3.8%; Score 81.4; DB 29; Length 733;  
Best Local Similarity 53.8%; Pred. No. 1.4e-09;  
Matches 189; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
Qy 1 GGATCCCAACTTTAGGAAGGATCTTAAATTTTAGTTATTAAGTTCAAGTTAGAAAA 60  
Db 603 GGTGGCCATCTCTACTTAAACATCTTAAACCTTGTTTTAAAGATCCTCAATTTGAAAT 544  
Qy 61 TCCTTACAGAGCTTTGAGTCCATGTATGATGATCCGTTGAAC-GGTGTACATGTCTCCG 119  
Db 543 TCCTTGGCTAAGGCTGCTAGATGATTTACTATATGTGTAATCTCTTTGCACTTTGA 484  
Qy 120 ATGGACTCACTGTGTTTCATTCGGAAAAGTTCGAAAGAGTGCATAAGAAATTTGATTTG 179  
Db 483 ATATTTTCATTTGCATTCATTTCTAAATAATTCATACTCATGAGTTAGTGCATTTATCCTA 424  
Qy 180 GATTTCTTCATCGTTGGTGGCTTCATGAGTGCTCAAGTGCTCCBATAATCAAAA 239  
Db 423 GATCTTTTAAACATCTGTAGTTCCCTCATGTGTTAATTAAGAGGTGTCACATTTTCCTTA 364  
Qy 240 GCGGAATCACAAATTTGAATGTGATTTGAATTCATTTTGTCTTAATGCAAAAAACAGGCA 299  
Db 363 GCATCTTACAAATTTGAACCTGAAATATTCATCCATTCCTAGGCGAGATGTTATTATG 304  
Qy 300 TCCATAGCTTTGTTTAAAGCAAAAACATCTCTCCGATTCATCCAT 350  
Db 303 TTTTGGCTTTTAAAGTTGTTATTTGACTCATTTTCTATCTTCTTCTCAGACCAT 253

RESULT 12  
CG817875/c  
LOCUS  
DEFINITION  
H11E21:MTp9H13, genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CG817875  
CG817875.1 GI:38273079

GSS.  
Glycine max (soybean)

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

1 (bases 1 to 872)

Shultz,J., Mekeem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,  
Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.

End sequencing of BACs comprising a provisional minimal tiling path  
from a fingerprint physical map of soybean (Glycine max) cultivar  
Forrest

Unpublished (2003)

Other GSSs: SOYDC43TH

Contact: Chris Town, J. L. Shultz and D. A. Lightfoot

The Center of Excellence in Soybean Research, Teaching and

Outreach, Southern Illinois University at Carbondale and Plant

Genomics, The Institute for Genomic Research

Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,

USA and 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 618 453 1797

Email: ga4082@siu.edu, jshultz@siu.edu, cdtown@tigr.org (URL:  
http://bioinformatics.siu.edu)

Clones approximating a minimum tiling path were re-arrayed from the

library master plates prior to sequencing.

For purposes of clone identification each clone name is a

concatenation of the original clone location and its new location

in the re-arrayed sequencing plates.

Seq primer: GTAATACGACTCACTATAGGCG

Class: BAC ends.

Location/Qualifiers

1..872

/organism="Glycine max"

/mol\_type="genomic DNA"

/cultivar="Forrest"

/db\_xref="taxon:3847"

/clone="H11E21:WTP9H13"

/clone\_lib="LargeInsertSoybeanGenLib"

/note="Organ: Leaves; Vector: pCLD04541 (pBELOBAC11 BcoRI  
clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
cv. Forrest seeds were grown in greenhouse for fourteen  
days. Nuclei were isolated and embedded in agarose,  
restriction digested with Hind III BamHI or EcoRI, large  
size DNA fragments were ligated in vector v41 (pCLD04541)  
and electro transformed in DH10a cells. About 90,000  
clones from BAC libraries were fingerprinted with HindIII  
and Hae III. Version 2 (automatic build) Contigs were  
built from 78,001 fingerprints. Contigs were manually  
examined to find the best non redundant tile path through  
the contigs representing about 13,000 clones. The clones  
were end sequenced."

FEATURES  
source

ORIGIN

Query Match 3.8%; Score 81.4; DB 29; Length 872;  
Best Local Similarity 53.8%; Pred. No. 1.4e-09;  
Matches 189; Conservative 0; Mismatches 161; Indels 1; Gaps 1;  
Qy 1 GGATCCCAACTTTTAGGAATGGATCTTAAATTTTAGTTATTAAGTTCAAGTTAGAAAA 60  
Db 596 GGTGGCCATCTCTACTTAAACATCTTAAACCTTGTTTTAAAGATCCTCAATTTGAAAT 537  
Qy 61 TCCTTACAGAGCTTTGAGTCCATGTATGATGATCCGTTGAACGGTGTACATGTCTCCG- 119  
Db 536 TCCTTGGCTAAGGCTGCTAGATGATTTACTATATGTGTAATCTCTTTTGCATATCTCGA 477  
Qy 120 ATGGACTCACTGTGTTTCATTCGGAAAAGTTTCGAAAGAGTGCATAAGAAATTTGATTTG 179  
Db 476 ATATTTTCATTTGCATTCATTTCTAAATAATTCATACTCATGAGTTAGTGCATTTATCCTA 417  
Qy 180 GATTTCTTCATCGCTGGTGGTGGCTTCATGAGTGCTCAAGAGTCCCTCAAGATATCAAAA 239

```

Db      416  GATCTTTTAAATCTCTAGTTCCTTCATGTTGTAATCGAAGAGTGTCACACATTTCTCTTA 357
QY      240  GCGAATCAAAATGGAATGATGAATCAATTTTGTCTTAATGCAACAAACAGGGCA 299
Db      356  GCACTCTTAAATTTGAAACCTTGAATATATTCATCCATTCCTAGGGCAGATGTTATTATG 297
QY      300  TTCAATAGCCCTTTGTTTAAAGCAAAACATCTTCTCGGATTCATCCCAT 350
Db      296  TTTTGGCTTTAAGTTGATTGTACTCGTTTCTATCTCTTCCTCGGACCAT 246

RESULT 13
LOCUS   CB340076                295 bp    mRNA    linear    EST 14-MAR-2003
DEFINITION   CA23EI02IRC_C04 Cabernet Sauvignon Leaf - CA23EI Vitis vinifera
ACCESSION   CB340076
VERSION     CB340076.1 GI:28960151
KEYWORDS     Vitis vinifera
SOURCE      Vitis vinifera
ORGANISM    Vitis vinifera
REFERENCE   1 (bases 1 to 295)
AUTHORS     Goes da Silva,F., Lim,H., Iandolino,A., Baek,J., Jones,K.,
            Walker,M.A. and Cook,D.R.
TITLE       Transcriptional responses of Vitis vinifera to infection by the
            bacterial pathogen Xylella fastidiosa
JOURNAL     Unpublished (2003)
COMMENT     Contact: Douglas Cook, PhD
            CAES Genome Facility
            UC Davis, Plant Pathology
            One Shields Ave, Davis, CA 95616, USA
            Tel: 530 754 6561
            Fax: 530 754 6617
            Email: drcook@ucdavis.edu
            Seq primer: GCCAAGCAATGGTCTAG.

FEATURES             source
     1..295
         /organism="Vitis vinifera"
         /mol_type="mRNA"
         /cultivar="Cabernet Sauvignon"
         /db_xref="taxon:29760"
         /clone="CA23EI02IRC_C04"
         /sex="Hermaphrodite"
         /dev_stage="Mid-season leaf material, collected July 25, 2001."
         /lab_host="DH5alpha"
         /clone_lib="Cabernet Sauvignon Leaf - CA23EI"
         /note="Organ: Leaf; Vector: pDNR; Site_1: Sfil; Site_2: Sfil; CA23EI is a cDNA library of Cabernet Sauvignon leaves. The leaves were collected on July 25, 2001, in Napa Valley, California, and represent leaves in mid-season development. These leaves were verified to be infected with the bacterial pathogen, Xylella fastidiosa, based on a diagnostic assay using PCR and Xylella-specific primer pairs. The plants were asymptomatic at the time of collection, but later developed symptoms. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
         5'-AAGCACTGTATCAACGACGATGGCATACGCGCGG-3' and
         5'-ATTTCTAGCGCGGACGCGGACATG-TT(30)NN-3'. Library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-3 kb size fraction."

ORIGIN
Query Match      3.8%; Score 81.2; DB 14; Length 295;
Best Local Similarity 63.8%; Pred. No. 1.4e-09;
Matches 139; Conservative 0; Mismatches 78; Indels 1; Gaps 1;

QY      64  TTACCAAGAGCTTTGAGTCCATTGATGACATCCGTTGAAC-GGTGFACATGTCCTCGATG 122

```

```

Db      76  TTTCCAAAGCTTGAATTCATTCACAAATCCCATGAACCTAAAAACATCTCTACAATA 135
QY      123  GACTCACTTGGTTTCATTCGAAAGCTTCGAAAGAGTGCATAAGATATTTGGAT 182
Db      136  GATTCAAGATCCTTTCATAAAAAAATAATTCATAATAATGCGCAAGGATGTTGATTCTAGAC 195
QY      183  TCTTTCACTCGGTTTGGTTCCTTCATGAGTGACCTCAAGAGTCTCTCAAAATATCAAAAGCC 242
Db      196  TCTTTAACTTGGTTGGTTCCTTCATGAGTAATTTCTAATAGTCTCCAAATCTCTTAGCC 255
QY      243  GAATCAAAAATTCGAAATGTGATTGAATTCATTTTCTC 280
Db      256  AATTTCATTTGAAACATGATTAATTCATTTCTATC 293

RESULT 14
LOCUS   CG816026/c              795 bp    DNA    linear    GSS 18-NOV-2003
DEFINITION   SOYFH38TV LargeInsertSoybeanGenLib Glycine max genomic clone
ACCESSION   CG816026
VERSION     CG816026.1 GI:38269500
KEYWORDS     Glycine max (soybean)
SOURCE      Glycine max
ORGANISM    Glycine max
REFERENCE   1 (bases 1 to 795)
AUTHORS     Shultz,J., Meksem,K., Shetty,J., Town,C.D., Koo,H., Potter,J.,
            Wakefield,K., Zhang,H., Wu,C. and Lightfoot,D.A.
TITLE       End sequencing of BACs comprising a provisional minimal tiling path
            from a fingerprint physical map of soybean (Glycine max) cultivar
            Forrest
JOURNAL     Unpublished (2003)
COMMENT     Contact: Chris Town, J. L. Shultz and D. A. Lightfoot
            The Center of Excellence in Soybean Research, Teaching and
            Outreach, Southern Illinois University at Carbondale and Plant
            Genomics, The Institute for Genomic Research
            Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62301-4415,
            USA and 9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 618 453 1797
            Email: ga4082@siu.edu, jlshultz@siu.edu, cdtown@tigr.org (URL:
            http://bioinformatics.siu.edu)
            Clones approximating a minimum tiling path were re-arrayed from the
            library master plates prior to sequencing.
            For purposes of clone identification each clone name is a
            concatenation of the original clone location and its new location
            in the re-arrayed sequencing plates.
            Seq primer: GTAATACGACTCACTATAGGGC
            Class: BAC ends.

FEATURES             Location/Qualifiers
     1..795
         /organism="Glycine max"
         /mol_type="genomic DNA"
         /cultivar="Forrest"
         /db_xref="taxon:3847"
         /clone="VH41L04:SEQ3H4"
         /note="Organ: Leaves; Vector: pCLD04541 (pBETOBAC11 EcorI
            clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)
            cv. Forrest seeds were grown in greenhouse for fourteen
            days. Nuclei were isolated and embedded in agarose,
            restriction digested with Hind III BamHI or EcorI, large
            size DNA fragments were ligated in vector V41 (pCLD04541)
            and electro transformed in DH10a cells. About 90,000
            clones from BAC libraries were fingerprinted with HindIII
            and Hae III. Version 2 (automatic build) Contigs were
            built from 78,001 fingerprints. Contigs were manually
            examined to find the best non redundant tile path through
            the contigs representing about 13,000 clones. The clones

```

were end sequenced."

/clone lib="LargeInsertSoybeanGenLib"  
/note="Organ: Leaves; Vector: pCLD04541 (pBeloBAC11 EcoRI  
clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
cv. Forrest seeds were grown in greenhouse for fourteen  
days. Nuclei were isolated and embedded in agarose,  
restriction digested with Hind III BamHI or EcoRI, large  
size DNA fragments were ligated in vector V41 (pCLD04541)  
and electro transformed in DH10a cells. About 90,000  
clones from BAC libraries were fingerprinted with HindIII  
and Hae III. Version 2 (automatic build) Contigs were  
built from 78,001 fingerprints. Contigs were manually  
examined to find the best non redundant tile path through  
the contigs representing about 13,000 clones. The clones  
were end sequenced."

## ORIGIN

Query Match 3.8%; Score 81.2; DB 29; Length 914;  
Best Local Similarity 53.3%; Pred. No. 1.5e-09;  
Matches 193; Conservative 0; Mismatches 168; Indels 1; Gaps 1;  
QY 1 GGATCCCACTTTTAGGAATGGATCTTAAATTTTATAGTTTCAAGTTAGAAAA 60  
DB 571 GGTGGCAATCTCTACTAGACATTTCAAACTTTTATTTATAGATCCTCATTTGGAAT 512  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTCATGACATCCGTGAAAC-GGTGTACATGTCTCCG 119  
DB 511 ATTTTTCCTAAAGATGCAAGATGATTAACATATGTGTGGAACCTCTTTTCATGTCTGT 452  
QY 120 ATGAGCTCACTTGGTTTCATTCGGAAAGTTGGAAGAGTGCATAGATATTCATTTTG 179  
DB 451 ATACTTTTCATTTGCATTCATCTCTAAATAGTTCATATTCATGAGTTAATGTGTTATCTTA 392  
QY 180 GATTCTTTTCACTCGGTTCGCTTCATGAGTGCACCTCAAGAGTTCCTCCAAATATCAAAA 239  
DB 391 GATCTTTTAACTCTGTGTGGCTTCATGTTTACTTTGAGGTATCCCAATATCTCTTT 332  
QY 240 GCCGAATCAAAATGAAATGTGATTAATTCATTTTGTCTAATGCACAAAAACAGGGCA 299  
DB 331 TCACCTTTTACAGTTTGATACCCCTAAAGTATTCATCCATTCCTAGGGCAGATGTAATATG 272  
QY 300 TCCATAGCCTTTGTGTTTAAAGCAAAACATTCCTTCCTCCGATTCATCCCATTCGCTCATC 359  
DB 271 TTTTTCGCTTTTAAAGTTGTATTTGATAGTCTTTCTTTCTTCTCCTCACTCCCAATCTTCCTTA 212  
QY 360 GG 361  
DB 211 GG 210

Search completed: September 19, 2004, 02:37:42  
Job time : 4975.1 secs

## ORIGIN

Query Match 3.8%; Score 81.2; DB 29; Length 795;  
Best Local Similarity 57.2%; Pred. No. 1.5e-09;  
Matches 166; Conservative 0; Mismatches 123; Indels 1; Gaps 1;  
QY 1 GGATCCCACTTTTAGGAATGGATCTTAAATTTTATAGTTTCAAGTTAGAAAA 60  
DB 603 GGTGGCAATCTCTCTAGACATCTTAATACCTTTGTTATGAGATCTTCATTTGAAAA 544  
QY 61 TCTTTACCAAGAGCTTTGAGTCCATTCATGACATCCGTGAAACGGTGTA-CATGTCTCCG 119  
DB 543 TCTTTCCCTAAGCTGCAAGATGATTTACTATATGTGTGACCTCTTTTCATGTCTTGA 484  
QY 120 ATGAGCTCACTTGGTTTCATTCGGAAAGTTGGAAGAGTGCATAGATATTCATTTTG 179  
DB 483 ATGCTTTTCATTTGCATTCATCTCTAAATAGATTCATATTCATGAGTTAATGTGTTATCTTA 424  
QY 180 GATTCTTTCACTCGGTTCGCTTCATGAGTGCACCTCAAGAGTTCCTCCAAATATCAAAA 239  
DB 423 GATCTTTTAACTCTGTGTGGCTTCATGTTTACTTTGAGGTGTCCTCCATTCCTTA 364  
QY 240 GCCGAATCAAAATGAAATGTGATTAATTCATTTTGTCTAATGCACA 289  
DB 363 GCATCTTCAAAATTTGAACCTTGAAATATTCATTCATTCCTCCAGGGCAGA 314

## RESULT 15

CG822083/c  
LOCUS 914 bp DNA linear GSS 18-NOV-2003  
DEFINITION SOYD15TV LargeInsertSoybeanGenLib Glycine max genomic clone  
H37B06:WTP13D6, genomic survey sequence.

CG822083  
CG822083.1 GI:38281410

## ACCESSION

VERSION

## KEYWORDS

SOURCE

## ORGANISM

Glycine max (soybean)

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

Glycine.

1 (bases 1 to 914)

Shultz J., Meksem, K., Shetty, J., Town, C.D., Koo, H., Potter, J.,

Wakefield, K., Zhang, H., Wu, C. and Lightfoot, D.A.

End sequencing of BACs comprising a provisional minimal tiling path

from a fingerprint physical map of soybean (Glycine max) cultivar

Forrest

Unpublished (2003)

Other GSSs: SOYD15TH

Contact: Chris Town, J. L. Shultz and D. A. Lightfoot

The Center of Excellence in Soybean Research, Teaching and

Outreach, Southern Illinois University at Carbondale and Plant

Genomics, The Institute for Genomic Research

Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,

USA and 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 618 453 1797

Email: gs409@siu.edu, jshultz@siu.edu, cdtown@igrr.org (URL:

http://bioinformatics.siu.edu)

Clones approximating a minimum tiling path were re-arrayed from the

library master plates prior to sequencing.

For purposes of clone identification each clone name is a

concatenation of the original clone location and its new location

in the re-arrayed sequencing plates.

Seq primer: GTAATACGACTCACTATAGGCG

Class: BAC ends.

Location/Qualifiers

1..914

/organism="Glycine max"

/mol\_type="genomic DNA"

/cultivar="Forrest"

/db\_xref="taxon:3847"

/clone="H37B06:WTP13D6"

## FEATURES

source